Least squares estimation for binary decision trees

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

dem Bereich Mathematik und Naturwissenschaften der
Technischen Universität Dresden

von

Dipl. Math. Nadine Albrecht
geboren am 17. 01. 1983 in Altdöbern

Eingereicht am: 29.04.2020
Tag der Disputation: 13.10.2020

1. Gutachter: Prof. Dr. Dietmar Ferger
2. Gutachter: Prof. Dr. Dr. h. c. Winfried Stute

Contents

1 Introduction

Decision tree learning is a data analytics method often used to extract information from data sets which are unstructured, too large or complex. The approach is characterized by the recursive partitioning of explanatory variables by a sequence of binary splits into segments and predicting the value of a response variable in each of these segments. This procedure then corresponds to a tree-structured model, a *binary decision tree*. The nodes of the tree indicate decisions, the branches the decisions choice and the leafs give the predicted value. The term CART (Classification and regression trees) is used to cover the large variety of algorithms to find optimal decision trees and was first introduced by Breiman et al. [Bre+84]. Here, Classification trees describe models with categorical input and regression trees with numerical input variables.

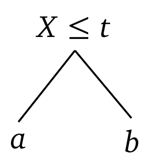
In this thesis we examine the case of a random vector $(X,Y) \in \mathbb{R}^2$ with the predictor variable X and response Y. We describe the relation between Y and X by a regression model

$$Y = m(X) + \epsilon\,,$$

with the conditional expectation $m(X) = \mathbb{E}(Y|X)$ and the error $\epsilon = Y - \mathbb{E}(Y|X)$, where $\mathbb{E}(Y^2) < \infty$. In what follows we use a binary regression tree as an approximation of the unknown regression curve $m(x) = \mathbb{E}(Y|X = x)$. Binary regression trees are then of the shape

$$g_{t,a,b}(x) = a\mathbb{1}_{(-\infty,t]}(x) + b\mathbb{1}_{(t,\infty)}(x)$$

and, therefore, characterized by the three parameters $(t,a,b) \in \mathbb{R}^3$. In our case $g_{t,a,b}(X)$ generates the simple tree



The parameter t is called *split point* and splits the feature space into two intervals $X \leq t$ and $X > t$. Whereas the constants a and b should predict Y on each side of the split point. The procedure of constructing such a simple prediction for Y is often used in situations in which an optimal distinction of the sample space of X or identification of a threshold is more important than the best possible prediction of Y. Now we assume that there exists a regression tree $g_{\tau,\alpha,\beta}$ with $\alpha \neq \beta$, which uniquely minimizes the L^2-distance from Y, i.e.

$$\mathbb{E}\left[\left(Y - g_{\tau,\alpha,\beta}(X)\right)^2\right] \leq \mathbb{E}\left[\left(Y - g_{t,a,b}(X)\right)^2\right]$$

for all $(t,a,b) \in \mathbb{R}^3$. For an independent and identically distributed (i. i. d.) sample $(X_i, Y_i)_{1\leq i\leq n}$ of (X,Y) we define the least squares estimator $(\tau_n, \alpha_n, \beta_n)$ for the parameters (τ, α, β) as a minimizer of the empirical process

$$S_n(t,a,b) := n^{-1} \sum_{i=1}^{n} \left(Y_i - a\mathbb{1}_{X_i \leq t} - b\mathbb{1}_{X_i > t}\right)^2 .$$

Estimators such as $(\tau_n, \alpha_n, \beta_n)$ are commonly called M-estimators. As a first asymptotic result we establish consistency of the estimator

$$(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$$

under the assumption that regression curve m and conditional variance $V(x) := \mathbb{E}(\epsilon^2|X = x)$ are bounded, X has a continuous distribution function F, and τ satisfies a 'well-separation' condition.

In order to derive distributional convergence, the vector

$$\theta_n = \left(\gamma_n^{(1)}(\tau_n - \tau), \gamma_n^{(2)}(\alpha_n - \alpha), \gamma_n^{(3)}(\beta_n - \beta)\right)$$

with the corresponding convergence rates $\gamma_n^{(1)}$, $\gamma_n^{(2)}$ and $\gamma_n^{(3)}$ is expressed as a minimizer (or maximizer) of an empirical process Z_n, which is a rescaled version of S_n, i.e.

$$\theta_n \in \mathrm{Argmin}(Z_n) .$$

There are different criteria, similar to the classical Continuous mapping Theorem (CMT), allowing to infer convergence of M-estimators from asymptotic results of Z_n. Thereby it is crucial whether the distribution of the limiting process allows for a single-element set of minimizers or not. For the case of single-element sets, ArgmaxCMT's are provided, e.g. by van der Vaart and Wellner [VW96, Theorem 3.2.2].

The asymptotic distribution of Z_n and the corresponding convergence rates in turn depend on the form of the unknown regression function in a neighborhood of the split point τ. For the special case of m being a regression tree itself, i.e.

$$m(x) = \alpha \mathbb{1}_{(-\infty,\tau]}(x) + \beta \mathbb{1}_{(\tau,\infty)}(x)\,,$$

m is called a *stump model*. Since the regression function and the best fitted regression tree then coincide, the setting is also a typical *change point model* and the parameter τ is called a *change point*. The authors Kosorok [Kos08, Section 14.5.1] and Seijo and Sen ([SS11a, section 5.1], [SS11b], [Sei12]) studied this model in the case where X and ϵ are assumed to be independent. The split point (or change point) estimator τ_n converges at a rate of n, whereas the estimators for the levels α_n and β_n converges at $\sqrt{n}$-rate. The limiting process of Z_n could be identified as a sum of three independent processes, a two-sided compound Poisson process and two random quadratic functions. A similar process occurs in the paper of Song, Banerjee and Kosorok [SBK16, Theorems 3.4 and 3.5], where the model $Y = m_n(X) + \epsilon$ varies at each stage n for smooth regression curves m_n and converges at various rates to a regression tree.

However, the compound Poisson process is piece-wise constant and possesses an interval of minimizers. One strategy used by the authors of [Kos08], [SS11a], [SS11b] and [SBK16] is to define a smallest-Argmax functional and derive a smallest-ArgmaxCMT, see [SS11a, Section 3]. Another approach for the non-unique case is pursued by Ferger [Fer04] for univariate and [Fer15] for multivariate processes. There the idea is to show classical convergence in distribution of Z_n to a process Z in the multivariate Skorokhod space and stochastic boundedness of θ_n. From this can be concluded that θ_n weakly converge to a so called Choquet capacity functional of the set of infimizers of Z. This functional is a generalizations of probability measures and suitable to construct confidence regions. We will utilize these results to generalize the stump model to regression functions of the form

$$m(x) = m_l(x)\mathbb{1}_{(-\infty,\tau]}(x) + m_r(x)\mathbb{1}_{(\tau,\infty)}(x)\,,$$

where $m_l(x)$ and $m_r(x)$ are continuous in a punctured neighborhood of the discontinuity point τ. The restriction that X and the error ϵ are independent will also be dropped.

Therefore, we find an appropriate decomposition of the rescaled process Z_n and show its weak convergence in the multivariate Skorokhod space. As one result it turns out that τ_n still converges at n-rate and α_n and β_n at $\sqrt{n}$-rates. The limiting process

$$Z(t,a,b) = 2(\beta-\alpha)Z^{(1)}(t) + Z^{(2)}(a) + Z^{(3)}(b)$$

is likewise a sum of a two-sided compound Poisson process, $2(\beta-\alpha)Z^{(1)}(t)$, and two random quadratic functions $Z^{(2)}(a)$ and $Z^{(3)}(b)$. The jump-size distributions for the two sides of $Z^{(1)}$ are derived by the conditional distributions of Y given X 'close' to the split point τ from the left and right side, respectively. The arrival-time distributions of $Z^{(1)}$ depend on the slope of F at τ. The processes $Z^{(2)}$ and $Z^{(3)}$ are derived by parameters depending on the distribution function F at τ, $\mathbb{E}(\mathbb{1}_{X\leq\tau}(\alpha-Y)^2)$ and $\mathbb{E}(\mathbb{1}_{X>\tau}(\beta-Y)^2)$, respectively.

To show stochastic boundedness of θ_n we derive numerous supremal inequalities for various empirical processes occuring by the investigation of the rescaled process Z_n. For this purpose, we generalize martingale and backwards martingale inequalities and use diverse martingale techniques like a Doob-Meyer decomposition.

The preliminary work used to proof this result also enables us to cover the continuous case. If m is continuous in a neighborhood of the split point, Banerjee and McKeague [BM07] pointed out that all three estimators $(\tau_n, \alpha_n, \beta_n)$ converge at a $n^{1/3}$-rate. Here the limiting process is driven by a two-sided Brownian motion with a quadratic drift function. We reproduce the result and, beyond that, observe that the assumptions made there have to be expanded. In order to ensure the convergence rate, the slope of the regression function m at the split point τ has to obey a certain condition. The limit process can be rearranged such that this condition is reflected in the parabolic drift function, which then opens in the correct direction in order to obtain an optimum.

The following five chapters are structured as follows. In Chapter 2 we give a precise description of the considered model and a formulation of the assumptions. In doing so we investigate the necessary condition about existence and uniqueness of the parameters (τ, α, β) in detail and state a simpler prescription for the calculation of $(\tau_n, \alpha_n, \beta_n)$. The mathematical foundation for multivariate Skorokhod spaces and the ArginfCMT are collected in Chapter 3. In Chapter 4 we derive the supremal inequalities. Chapter 5 is devoted to show the consistency of the estimator. Thereby we can use results from Chapter 4 and proof a generalization of Chang's theorem [SW86, p. 424]. Chapter 6 is intended to apply the Arginf-CMT for the case where the split point is a discontinuity point of m. To that we introduce the rescaled process and proof its weak convergence. To show the stochastic boundedness condition we use the inequalities of Chapter 4. Those results then allow us to show how to build confidence regions. In Chapter 7 we similarly proceed to apply the Argmax-CMT for the continuous case.

2 Estimation of binary decision trees

2.1 Setting and estimator

First we give a detailed description of the model and introduce notations which are valid throughout this work.

Let (X, Y) be a vector of real valued random variables defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$. X is the predictor and Y, influenced by X, is the response variable. This dependence structure is described by the joint distribution of (X, Y)

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_{Y|X}$$

(in the sense of [GS77, Proposition 1.8.10 together with Remark 5.3.9]), where $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$ is the distribution of X and $\mathbb{P}_{Y|X}$ is a regular version of the conditional distribution of Y given X. Denote by $F(x) := \mathbb{P}_X((-\infty, x])$ the corresponding right continuous distribution function attached to X and assume that $\mathbb{E}Y^2 < \infty$. Since Y is integrable the conditional expectation $m(X) := \mathbb{E}(Y|X)$ exists. A fundamental property of m is that it is the best prediction for Y by virtue of the information of $\sigma(X)$, in the sense that m minimizes the L^2-distance from Y (see [Kle06, Corollary 8.16]. The random error caused by the use of this model is denoted by $\epsilon := Y - m(X)$. By basic properties of conditional expectations, then $\mathbb{E}(\epsilon|X) = 0$.

We consider a non-parametric regression setting in which m is supposed to be a priori unknown. The existence of the regression function $m(x) := \mathbb{E}(Y|X = x)$ is established by the factorization lemma ([Kle06, Corollary 1.97]). The relation between $\mathbb{P}_{Y|X}$ and m then reads

$$m(x) = \int y \, \mathbb{P}_{Y|X}(x, \mathrm{d}y)$$

for $\mathbb{P}_X$-almost all $x \in \mathbb{R}$ (see [GS77, Proposition 5.3.12]). Thus m is $\mathbb{P}_X$-almost surely uniquely determined. Furthermore, $V(x) := \mathbb{E}(\epsilon^2|X = x)$ denotes the conditional variance of Y given $X = x$.

In this thesis we use a binary decision tree

$$g_{t,a,b}(x) = a\mathbb{1}_{(-\infty,t]}(x) + b\mathbb{1}_{(t,\infty)}(x)\,,$$

which is determined by the three parameters $(t, a, b) \in \mathbb{R}^3$ as a model for the unknown function m. First we formulate the inevitable conditions of existence and uniqueness for the best L^2-approximation of Y to a decision tree.

(A1) $S(t, a, b) := \mathbb{E}\big(Y - a\mathbb{1}_{X\le t} - b\mathbb{1}_{X>t}\big)^2$ has a unique minimizing set of parameters $(\tau, \alpha, \beta) \in \mathbb{R}^3$, where $\alpha \neq \beta$.

Furthermore we need some additional assumptions:

(A2) There exists an ε-neighborhood $U_\varepsilon(\tau)$ of τ and constants $0 < \underline{L} < \overline{L} < \infty$ such that F is continuous in $U_\varepsilon(\tau)$ with

$$\underline{L}|u - v| \le |F(u) - F(v)| \le \overline{L}|u - v|$$

for all $u, v \in U_\varepsilon(\tau)$. In addition, the right- and left-hand derivatives, $F'_+(\tau)$ and $F'_-(\tau)$, exist.

(A3) V is continuous and bounded in a punctured ε-neighbourhood $U_\varepsilon(\tau) \setminus \{\tau\}$.

Now let $(X_i, Y_i)_{1\le i\le n}$ be an n-sized i. i. d. sample of (X, Y) with the corresponding empirical distribution $Q_n := n^{-1}\sum_{i=1}^n \delta_{(X_i,Y_i)}$, where $\delta_{(x,y)}$ is Dirac's measure at the point (x, y). Then we introduce the empirical equivalent to $S(t, a, b)$ by

$$\begin{aligned} S_n(t, a, b) &:= \int_{\mathbb{R}^2} \big(y - a\mathbb{1}_{x\le t} - b\mathbb{1}_{x>t}\big)^2 Q_n(\,\mathrm{d}(x, y)) \\ &= n^{-1}\sum_{i=1}^n \big(Y_i - a\mathbb{1}_{X_i\le t} - b\mathbb{1}_{X_i>t}\big)^2 \end{aligned} \tag{2.1}$$

and define an M-estimator of the parameters (τ, α, β) by a measurable choice

$$(\tau_n, \alpha_n, \beta_n) \in \operatorname*{Argmin}_{(t,a,b)\in R^3} S_n(t, a, b)\,. \tag{2.2}$$

For more convenience we introduce the following expectations

$$
\begin{aligned}
&\bar{F}(t) := \mathbb{E}(\mathbb{1}_{X>t}) \\
&H(t) := \mathbb{E}(Y\mathbb{1}_{X\leq t}) = \mathbb{E}\left(\mathbb{E}\left(Y\mathbb{1}_{X\leq t}|X\right)\right) = \mathbb{E}\left(\mathbb{1}_{X\leq t}m(X)\right) = \int_{(-\infty,t]} m(x)F(\mathrm{d}x) \\
&\bar{H}(t) := \mathbb{E}(Y\mathbb{1}_{X>t}) = \int_{(t,\infty)} m(x)F(\mathrm{d}x)
\end{aligned}
\tag{2.3}
$$

and their corresponding empirical processes

$$
\begin{aligned}
&F_n(t) := n^{-1}\sum_{i=1}^{n}\mathbb{1}_{X_i\leq t}\,, \quad \bar{F}_n(t) := n^{-1}\sum_{i=1}^{n}\mathbb{1}_{X_i>t}\,, \\
&H_n(t) := n^{-1}\sum_{i=1}^{n}Y_i\mathbb{1}_{X_i\leq t}\,, \quad \bar{H}_n(t) := n^{-1}\sum_{i=1}^{n}Y_i\mathbb{1}_{X_i>t}\,.
\end{aligned}
$$

We denote the set where the distribution function F is neither zero nor one by

$$
T_F := \{x\in\mathbb{R}\,; 0 < F(x) < 1\}\ .
$$

2.2 Assumption $(A1)$

The existence and uniqueness condition (A1) is a strong requirement which implies several properties. These will be investigated in detail now, whereas the normal equations have already been mentioned in [BM07, p. 546]

Lemma 2.1 If condition **(A1)** holds, then

(i) $\tau \in T_F$.

(ii) For each $t \in T_F$ the set $\operatorname{Argmin}_{(a,b)\in\mathbb{R}^2} S(t,a,b)$ is a singleton and

$$
(a(t),b(t)) := \operatorname*{argmin}_{(a,b)\in\mathbb{R}^2} S(t,a,b) = \left(\frac{H(t)}{F(t)},\frac{\bar{H}(t)}{\bar{F}(t)}\right) = (\mathbb{E}(Y|X\leq t),\mathbb{E}(Y|X>t))\ .
$$

(iii) (α,β) satisfies the normal equations

$$
\alpha = \mathbb{E}(Y|X\leq\tau) \quad \text{and} \quad \beta = \mathbb{E}(Y|X>\tau)\,.
$$

As a result

$$\mathbb{E}\left(\mathbb{1}_{X\le\tau}(Y-\alpha)\right)=0 \quad\text{and}\quad \mathbb{E}\left(\mathbb{1}_{X>\tau}(Y-\beta)\right)=0\,.$$

(iv) The split point τ can be characterized by $\tau=\operatorname{argmax}_{t\in\mathbb{R}} M(t)$, where

$$M(t):=\begin{cases}\frac{H^2(t)}{F(t)}+\frac{\bar{H}^2(t)}{\bar{F}(t)} & \text{if } t\in T_F\\ (\mathbb{E}Y)^2 & \text{if } t\notin T_F\end{cases}.$$

(v) If $m(x)$ is continuous in a punctured ε-neighborhood $U_\varepsilon(\tau)\setminus\{\tau\}$ of τ, where τ is an interior point of T_F, then

$$2(\beta-\alpha)\left(m(\tau+)-\frac{\alpha+\beta}{2}\right)\ge 0 \quad\text{and}\quad 2(\beta-\alpha)\left(\frac{\alpha+\beta}{2}-m(\tau-)\right)\ge 0\,.$$

(vi) If $m(x)$ is continuous in an ε-neighborhood $U_\varepsilon(\tau)$ of τ, then

$$m(\tau)=\frac{\alpha+\beta}{2}\,.$$

(vii) If $m(x)$ is once and $F(x)$ twice differentiable in a neighborhood $U_\varepsilon(\tau)$ with $F'(\tau)>0$, then

$$\frac{m'(\tau)}{\beta-\alpha}\ge\frac{1}{4}\frac{F'(\tau)}{F(\tau)-F(\tau)^2}\,. \tag{2.4}$$

Proof. (i) First compute

$$\begin{aligned}S(t,a,b)&=\mathbb{E}(Y^2)+a^2\mathbb{E}(\mathbb{1}_{\{X\le t\}})+b^2\mathbb{E}(\mathbb{1}_{\{X>t\}})-2a\mathbb{E}(Y\mathbb{1}_{\{X\le t\}})-2b\mathbb{E}(Y\mathbb{1}_{\{X>t\}})\\ &=\begin{cases}\mathbb{E}(Y-b)^2 & F(t)=0\\ \mathbb{E}(Y^2)+a^2F(t)+b^2\bar{F}(t)-2aH(t)-2b\bar{H}(t) & t\in T_F\\ \mathbb{E}(Y-a)^2 & F(t)=1\,.\end{cases}\end{aligned} \tag{2.5}$$

To obtain a contradiction, suppose that $\tau\notin T_F$. Then either $F(\tau)=0$ or $F(\tau)=1$, and $S(\tau,a,b)$ would be constant in the argument a or b, respectively; a contradiction to the uniqueness condition of α and β.

(ii) Fix $t \in T_F$, then the partial derivatives of S in a and b exist and

$$(\partial_a S, \partial_b S) = \left(2aF(t) - 2H(t), 2b\bar{F}(t) - 2\bar{H}(t)\right) .$$

While considering the first-order necessary condition for a minimum in S, $(\partial_a S, \partial_b S) = (0,0)$, we get

$$a(t) = \frac{H(t)}{F(t)} \quad \text{and} \quad b(t) = \frac{\bar{H}(t)}{\bar{F}(t)} .$$

The corresponding Hessian matrix H_S, being of the form

$$\mathrm{H}_S = \left(\partial_x \partial_y\, S\right)_{x,y \in \{a,b\}} = \begin{pmatrix} 2F(t) & 0 \\ 0 & 2\bar{F}(t) \end{pmatrix} ,$$

is positive definite. Hence, $S(t,a,b)$ is strictly convex in (a,b) and $\operatorname{argmin}_{(a,b)\in R^2} S(t,a,b)$ is a singleton. Moreover, with the definition of elementary conditional expectations one gets

$$a(t) = \mathbb{E}(Y|X \leq t) \quad \text{and} \quad b(t) = \mathbb{E}(Y|X > t) .$$

(iii) A standard rule for interchanging the order of minimization, see [RW98, Proposition 1.35], gives the following specification of the, as required, singleton set of minimizing arguments

$$\begin{aligned} &\operatorname*{argmin}_{(t,a,b)\in\mathbb{R}^3} S(t,a,b) \\ &\quad = \left\{(\tau,\alpha,\beta);\ \tau \in \operatorname*{argmin}_{t\in\mathbb{R}} \left(\inf_{(a,b)\in\mathbb{R}^2} S(t,a,b)\right) \text{ and } (\alpha,\beta) \in \operatorname*{argmin}_{(a,b)\in\mathbb{R}^2} S(\tau,a,b)\right\} . \end{aligned} \tag{2.6}$$

By (i) and (ii) we obtain $(\alpha,\beta) = (a(\tau), b(\tau)) = (\mathbb{E}(Y|X \leq \tau), \mathbb{E}(Y|X > \tau))$. Therefore,

$$\mathbb{E}\left(\mathbb{1}_{X\leq\tau}(Y-\alpha)\right) = \mathbb{E}\left(Y\mathbb{1}_{X\leq\tau}\right) - \mathbb{E}(Y|X \leq \tau)\mathbb{E}\left(\mathbb{1}_{X\leq\tau}\right) = 0$$

and, accordingly, $\mathbb{E}\left(\mathbb{1}_{X>\tau}(Y-\beta)\right) = 0$.

(iv) Use Equation (2.5) and (ii) to see that

$$\inf_{(a,b)\in\mathbb{R}^2} S(t,a,b) = \begin{cases} \mathbb{E}Y^2 - \frac{H^2(t)}{F(t)} - \frac{\bar{H}^2(t)}{\bar{F}(t)} & t \in T_F \\ \mathbb{E}Y^2 - (\mathbb{E}Y)^2 & t \notin T_F \,. \end{cases}$$

By (i) and Equation (2.6) then $\tau = \operatorname{argmin}_{t\in T_F} \inf_{(a,b)\in\mathbb{R}^2} S(t,a,b)$, whereas this is true if and only if $\tau = \operatorname{argmax}_{t\in T_F} M(t)$.

(v) For each $v \neq 0$ with $(\tau - v, \tau + v) \in T_F$ condition (A1) implies the following strict inequality

$$0 < S(\tau+v,\alpha,\beta) - S(\tau,\alpha,\beta) = \begin{cases} -2(\alpha-\beta) \int\limits_{(\tau,\tau+v]} \left(m(x) - \frac{\alpha+\beta}{2}\right) F(\mathrm{d}x) & v > 0 \\ 2(\alpha-\beta) \int\limits_{(\tau+v,\tau]} \left(m(x) - \frac{\alpha+\beta}{2}\right) F(\mathrm{d}x) & v < 0 \end{cases} . \tag{2.7}$$

Without loss of generality, assume $\alpha < \beta$. The case where $\alpha > \beta$ can be proved in much the same way. Let $(v_n)_{n\in\mathbb{N}} \subseteq (0,\varepsilon)$ be a sequence with $v_n \downarrow 0$. For each $n \in \mathbb{N}$ we can write (2.7) as

$$\int\limits_{(\tau-v_n,\tau]} \left(m(x) - \frac{\alpha+\beta}{2}\right) F(\mathrm{d}x) < 0 < \int\limits_{(\tau,\tau+v_n]} \left(m(x) - \frac{\alpha+\beta}{2}\right) F(\mathrm{d}x) \,.$$

By the continuity of m in $U_\varepsilon(\tau) \setminus \{\tau\}$ and the mean value theorem for Lebesgue-Stieltjes integrals, see [HS75, Theorem 21.69], there exist sequences $(\xi_n)_{n\in\mathbb{N}}$ and $(\tilde{\xi}_n)_{n\in\mathbb{N}}$ with $\xi_n \in (\tau, \tau + v_n)$ and $\tilde{\xi}_n \in (\tau - v_n, \tau)$ for each $n \in \mathbb{N}$, such that

$$\begin{aligned} 0 &< \int\limits_{(\tau,\tau+v_n]} \left(m(x) - \frac{\alpha+\beta}{2}\right) F(\mathrm{d}x) = \left(m(\xi_n) - \frac{\alpha+\beta}{2}\right)(F(\tau+v_n) - F(\tau)) \,, \\ 0 &> \int\limits_{(\tau-v_n,\tau]} \left(m(x) - \frac{\alpha+\beta}{2}\right) F(\mathrm{d}x) = \left(m(\tilde{\xi}_n) - \frac{\alpha+\beta}{2}\right)(F(\tau) - F(\tau-v_n)) \,. \end{aligned} \tag{2.8}$$

These inequalities lead to

$$m(\tilde{\xi}_n) < \frac{\alpha+\beta}{2} < m(\xi_n) \,.$$

Finally, since $\xi_n \searrow \tau$ and $\tilde{\xi}_n \nearrow \tau$, as $n \to \infty$, the inequalities are proved.

(vi) Follows from (v) if we set $m(\tau-) = m(\tau+)$.

(vii) By (i),(ii) and (iv) we know that $\tau = \operatorname{argmax}_{t \in T_F} M(t)$, where

$$M(t) = a(t)H(t) + b(t)\bar{H}(t)\,.$$

For each $t \in U_\varepsilon(\tau)$ we apply A.5 to obtain

$$H'(t) = \lim_{h \to 0} \frac{H(t+h) - H(t)}{F(t+h) - F(t)} \frac{F(t+h) - F(t)}{h} = m(t)F'(t)\,. \tag{2.9}$$

With the same argument

$$\bar{H}'(t) = -m(t)F'(t)\,, \quad a'(t) = F'(t)\frac{m(t) - a(t)}{F(t)}\,, \quad b'(t) = F'(t)\frac{b(t) - m(t)}{\bar{F}(t)}$$

and, thus,

$$\begin{aligned} M'(t) =& F'(t)\big[2m(t)(a(t) - b(t)) + b^2(t) - a^2(t)\big] \\ M''(t) =& F''(t)\big[2m(t)(a(t) - b(t)) + b^2(t) - a^2(t)\big] \\ &+ 2F'(t)\big[m'(t)(a(t) - b(t)) + a'(t)(m(t) - a(t)) + b'(t)(b(t) - m(t))\big]\,. \end{aligned}$$

With $F'(\tau) > 0$ and the first-order necessary condition for τ to be an optimum in M, that is $M'(\tau) = 0$, then

$$M''(\tau) = F'(\tau)\frac{(\alpha - \beta)^2}{2}\left[\frac{4m'(\tau)}{\alpha - \beta} + \frac{F'(\tau)}{F(\tau) - F^2(\tau)}\right]\,.$$

Suppose, contrary to our claim, that $\frac{m'(\tau)}{\beta - \alpha} < \frac{1}{4}\frac{F'(\tau)}{F(\tau) - F(\tau)^2}$, then $M''(\tau) > 0$, which is sufficient for $M(\tau)$ to be a local minimum; a contradiction to (iv). □

As already mentioned in Section 2.1, $m(X)$ is that $\sigma(X)$-measurable random variable which minimizes the L^2-distance from Y. Using a decision tree (or any other square-integrable function) g instead of m causes an error which we may characterize as follows.

For each $B \in \mathscr{B}(\mathbb{R})$ we have

$$
\begin{aligned}
&\mathbb{E}\left(\mathbb{1}_{X^{-1}(B)}(Y-g(X))^2\right) \\
&= \int\limits_{X^{-1}(B)} (Y-m(X))^2 \, \mathrm{d}\mathbb{P} + 2\int\limits_{X^{-1}(B)} (Y-m(X))(m(X)-g(X)) \, \mathrm{d}\mathbb{P} + \int\limits_{X^{-1}(B)} (m(X)-g(X))^2 \, \mathrm{d}\mathbb{P} \\
&= \int\limits_B \int (y-m(x))^2 \, \mathbb{P}_{Y|X}(x, \mathrm{d}y) F(\mathrm{d}x) \\
&\qquad\qquad + \int\limits_B (m(x)-g(x)) \int y - m(x) \, \mathbb{P}_{Y|X}(x, \mathrm{d}y) F(\mathrm{d}x) \\
&\qquad\qquad\qquad + \int\limits_B (m(x)-g(x))^2 \, F(\mathrm{d}x) \\
&= \int\limits_B V(x) F(\mathrm{d}x) + \int\limits_B (m(x)-g(x))^2 \, F(\mathrm{d}x) \, . \qquad (2.10)
\end{aligned}
$$

If $B = \mathbb{R}$ we observe that the error can be decomposed into a sum of two errors. The first one is the conditional variance averaged over X. It is independent of g and, hence, a lower bound for the error. The second term is an average of the error caused by the use of g as a model function for the regression curve m.

2.3 Calculating the estimator

Let us denote the sampling interval by

$$T_{F_n} = \{x \in \mathbb{R}\,; 0 < F_n(x) < 1\} \; .$$

By $X_{i:n}$ we will denote the i-th order statistic of the sample $X_1, \ldots, X_n$. Set

$$\bar{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i \, .$$

Within this section we adhere to the convention that $0/0 = 0$. The following lemma enables us to compute $(\tau_n, \alpha_n, \beta_n)$.

Lemma 2.2 For each $(\tau_n, \alpha_n, \beta_n) \in \operatorname*{Argmin}_{(t,a,b) \in R^3} S_n(t, a, b)$ we have

(i) $\tau_n \in \operatorname*{Argmax}_{t\in\mathbb{R}} M_n(t)$, where

$$M_n(t) := \begin{cases} \frac{H_n^2(t)}{F_n(t)} + \frac{\bar{H}_n^2(t)}{\bar{F}_n(t)} & \text{if } t \in T_{F_n} \\ \bar{Y}_n^2 & \text{if } t \notin T_{F_n} \end{cases}, \tag{2.11}$$

(ii) and $(\alpha_n, \beta_n) = (a_n(\tau_n), b_n(\tau_n))$, where

$$(a_n(t), b_n(t)) := \left(\frac{H_n(t)}{F_n(t)}, \frac{\bar{H}_n(t)}{\bar{F}_n(t)} \right), \quad t \in \mathbb{R}.$$

Proof. (i) Similar to the proof of Lemma 2.1(i) and (iv) we can show that

$$S_n(t,a,b) = \begin{cases} n^{-1}\sum_{i=1}^n (Y_i - b)^2 & t < X_{1:n} \\ n^{-1}\sum_{i=1}^n Y_i^2 + a^2 F_n(t) + b^2 \bar{F}_n(t) - 2aH_n(t) - 2b\bar{H}_n(t) & t \in T_{F_n} \\ n^{-1}\sum_{i=1}^n (Y_i - a)^2 & t \geq X_{n:n}. \end{cases} \tag{2.12}$$

and

$$\inf_{(a,b)\in\mathbb{R}^2} S_n(t,a,b) = \begin{cases} n^{-1}\sum_{i=1}^n Y_i^2 - \frac{H_n^2(t)}{F_n(t)} - \frac{\bar{H}_n^2(t)}{\bar{F}_n(t)} & t \in T_{F_n} \\ n^{-1}\sum_{i=1}^n Y_i^2 - \bar{Y}_n^2 & t \notin T_{F_n} \end{cases}.$$

Likewise, with the rule for interchanging the order of minimization (see [RW98, Proposition 1.35]), for each $n \in \mathbb{N}$ we have

$$\begin{aligned} &\operatorname*{Argmin}_{(t,a,b)\in\mathbb{R}^3} S_n(t,a,b) \\ &= \left\{ (\tau_n, \alpha_n, \beta_n);\ \tau_n \in \operatorname*{Argmin}_{t\in\mathbb{R}} \left(\inf_{(a,b)\in\mathbb{R}^2} S_n(t,a,b) \right) \text{ and } (\alpha_n,\beta_n) \in \operatorname*{Argmin}_{(a,b)\in\mathbb{R}^2} S_n(\tau_n, a, b) \right\} \end{aligned}$$

and, hence

$$\tau_n \in \operatorname*{Argmin}_{t\in\mathbb{R}} \left(\inf_{(a,b)\in\mathbb{R}^2} S_n(t,a,b) \right) = \operatorname*{Argmin}_{t\in\mathbb{R}} (-M_n(t)) = \operatorname*{Argmax}_{t\in\mathbb{R}} M_n(t) \qquad \square$$

Finally we will examine the empirical process S_n from (2.1) in order to find suitable decompositions for the rescaled versions of S_n. To that end, note that

$$S_n(t, a, b) = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i \le t} (Y_i - a)^2 + n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i > t} (Y_i - b)^2 \,.$$

Furthermore,

$$\begin{aligned}
&S_n(\tau + t, \alpha, \beta) - S_n(\tau, \alpha, \beta) \\
&= n^{-1} \sum_{i=1}^{n} \left(\mathbb{1}_{X_i \le \tau + t} - \mathbb{1}_{X_i \le \tau} \right) \left[(Y_i - \alpha)^2 - (Y_i - \beta)^2 \right] \\
&= 2(\beta - \alpha) n^{-1} \sum_{i=1}^{n} \left(\mathbb{1}_{X_i \le \tau + t} - \mathbb{1}_{X_i \le \tau} \right) \left(Y_i - \frac{\alpha + \beta}{2} \right)
\end{aligned}$$

and

$$\begin{aligned}
&S_n(\tau + t, \alpha + a, \beta + b) - S_n(\tau + t, \alpha, \beta) \\
&= n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i \le \tau + t} \left(a^2 - 2a(Y_i - \alpha) \right) + n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i > \tau + t} \left(b^2 - 2b(Y_i - \beta) \right) .
\end{aligned}$$

The empirical process S_n centered at the point (τ, α, β) can be decomposed

$$\begin{aligned}
&S_n(\tau + t, \alpha + a, \beta + b) - S_n(\tau, \alpha, \beta) \\
&= [S_n(\tau + t, \alpha, \beta) - S_n(\tau, \alpha, \beta)] + [S_n(\tau + t, \alpha + b, \beta + b) - S_n(\tau + t, \alpha, \beta)] \\
&= 2(\beta - \alpha) n^{-1} \sum_{i=1}^{n} \left(\mathbb{1}_{X_i \le \tau + t} - \mathbb{1}_{X_i \le \tau} \right) \left(Y_i - \frac{\alpha + \beta}{2} \right) \\
&\qquad\qquad - 2a n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i \le \tau + t} (Y_i - \alpha) - 2b n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i > \tau + t} (Y_i - \beta) \\
&\qquad\qquad\qquad + a^2 n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i \le \tau + t} + b^2 n^{-1} \sum_{i=1}^{n} \mathbb{1}_{X_i > \tau + t}
\end{aligned}$$

$$
\begin{aligned}
&= 2(\beta-\alpha)n^{-1}\sum_{i=1}^{n}\left(\mathbb{1}_{X_i\le\tau+t}-\mathbb{1}_{X_i\le\tau}\right)\left(Y_i-\frac{\alpha+\beta}{2}\right)\\
&\qquad\qquad\qquad + 2an^{-1}\sum_{i=1}^{n}\mathbb{1}_{X_i\le\tau}(\alpha-Y_i)+2bn^{-1}\sum_{i=1}^{n}\mathbb{1}_{X_i>\tau}(\beta-Y_i)\\
&\qquad\qquad\qquad + 2n^{-1}\sum_{i=1}^{n}\left(\mathbb{1}_{X_i\le\tau+t}-\mathbb{1}_{X_i\le\tau}\right)[(b-a)Y_i+(a\alpha-b\beta)]\\
&\qquad\qquad\qquad\qquad + a^2n^{-1}\sum_{i=1}^{n}\mathbb{1}_{X_i\le\tau+t}+b^2n^{-1}\sum_{i=1}^{n}\mathbb{1}_{X_i>\tau+t}\,. \qquad (2.13)
\end{aligned}
$$

3 Multivariate cadlag processes

In this chapter the mathematical foundations are gathered from the literature and completed by further results.

In Section 3.1 we recall the definition of the multivariate Skorokhod space $D(\mathbb{R}^d)$. We give a characterization of convergence with respect to the Skrokhod topology. Furthermore, we introduce the Skorokhod product space and present criteria for the convergence with respect to the product topology. In Section 3.2 we recall the Arginf-functional of a process X in $D(\mathbb{R}^d)$ and provide a CMT for convergence almost surely and in probability. Finally, we formulate a corresponding CMT concerning the generalized convergence in distribution.

3.1 The multivariate Skorokhod space

We use the definition of the multivariate Skorokhod space $D(\mathbb{R}^d)$ from [Fer15, p. 13]. It extends the definition of the function space $D([0,\infty)^d)$ from [ŁR86, p. 329].

Definition 3.1 (Multivariate Skorokhod space) Assume that $d \in \mathbb{N}$ and R_k, $1 \le k \le d$, is one of the relations $<$ or $\geq$. For $R = (R_1, \ldots, R_d)$ and all $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$ then

$$Q_R(t) := \left\{ s \in \mathbb{R}^d;\, s_k R_k t_k,\, 1 \le k \le d \right\}$$

is called the *R-quadrant of* t. For a function $f : \mathbb{R}^d \to \mathbb{R}$ the limit (if it exists)

$$f_{Q_R}(t) := \lim_{Q_R \ni s \to t} f(s)$$

is called the *R-quadrant-limit of* f *in* t. The *multivariate Skorokhod space* $D(\mathbb{R}^d)$ is the set of all functions $f : \mathbb{R}^d \to \mathbb{R}$ which satisfy for each $t \in \mathbb{R}^d$ the conditions

(i) $f_{Q_R}(t)$ exists for all of the 2^d R-quadrants $Q_R(t)$,

(ii) $f_{Q_R}(t) = f(t)$ for $R = (\geq, \ldots, \geq)$.

Suppose that $I^d := [-a, a] \subseteq \mathbb{R}^d$ for some $0 < a = (a_1, \ldots, a_d)$ then the subspace $D(I^d)$ denotes the set of all restrictions of functions $f \in D(\mathbb{R}^d)$ to I^d.

Based on the notion for the univariate case $D(\mathbb{R})$, the functions in $D(\mathbb{R}^d)$ are also called 'right-continuous with left limits (càdlàg)'. Now we will show that the empirical process S_n from (2.1) has càdlàg trajectories.

Lemma 3.2 For each $n \in \mathbb{N}$, the trajectories $(t, a, b) \mapsto S_n(\omega, t, a, b)$ are functions in $D(\mathbb{R}^3)$ for all $\omega \in \Omega$.

Proof. Let $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ and $Q_R(x) = Q_{(R_1,R_2,R_3)}(x_1, x_2, x_3)$ be an R-Quadrant of x. Let $(x^{(k)})_{k\in\mathbb{N}} \subseteq Q_R$ be a sequence with $\lim_{k\to\infty} x^{(k)} = x$. Since convergence with respect to a norm in $\mathbb{R}^3$ is equivalent to component-wise convergence ([Heu91, Proposition 109.8]) and $Q_R(x) = Q_{R_1}(x_1) \times Q_{R_2}(x_2) \times Q_{R_3}(x_3)$, this is equivalent to $x_i^{(k)} \to x^{(k)}$ in Q_{R_i}, $1 \le i \le 3$. With the representation of S_n from Equation (2.12) we have that

$$\begin{aligned} S_{Q_R}(x) &= \lim_{\substack{x_i^{(k)} \to x^{(k)} \\ 1\le i\le 3}} S_n(x_1^{(k)}, x_2^{(k)}, x_3^{(k)}) \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 \\ &\quad + \lim_{\substack{x_i^{(k)} \to x^{(k)} \\ 1\le i\le 3}} \left[(x_2^{(k)})^2 F_n(x_1^{(k)}) + (x_3^{(k)})^2 \bar{F}_n(x_1^{(k)}) - 2x_2^{(k)} H_n(x_1^{(k)}) - 2x_3^{(k)} \bar{H}_n(x_1^{(k)}) \right] . \end{aligned}$$

We know that $t \mapsto \mathbb{1}_{X_i \le t}$ is càdlàg and thus F_n, $\bar{F}_n$, H_n and $\bar{H}_n$ are too. Now conditions (i) and (ii) from Definition 3.1 follows from the fact, that the maps $(x, y) \mapsto x + y$ and $(x, y) \mapsto xy$ are continuous ([For08, Proposition 2.7]). $\square$

Similar to the univariate case in [Bil99, Section 12] one can define the Skorokhod metric s_d for the multivariate case. This metric generates a topology which is called Skorokhod topology and which makes $D(\mathbb{R}^d)$ a separable space. Moreover, there exists another metric which also generates the Skorokhod topology but beyond that lets $D(\mathbb{R}^d)$ become separable and complete (see for instance [ŁR86, p. 332]). Now we give a characterization of convergence in $D(\mathbb{R}^d)$ induced by the Skorokhod topology.

Lemma 3.3 Let Λ_d be the set of all transformations $\lambda : \mathbb{R}^d \to \mathbb{R}^d$ of the form $\lambda(t_1, \ldots, t_d) = (\lambda_{(1)}(t_1), \ldots, \lambda_{(d)}(t_d))$, where each $\lambda_{(i)} : \mathbb{R} \to \mathbb{R}$, $1 \le i \le d$, is continuous and strictly increasing with $\lambda_{(i)}(-\infty) = -\infty$ and $\lambda_{(i)}(\infty) = \infty$. If f and the sequence $(f_n)_{n\in\mathbb{N}}$ are functions in $D(\mathbb{R}^d)$ then f_n converges to f with respect to s_d (for short, $f_n \xrightarrow{s_d} f$) if and only if there exists a sequence $(\lambda_n)_{n\in\mathbb{N}} \subseteq \Lambda_d$ such that

(i) $\sup_{t\in\mathbb{R}^d} \|\lambda_n(t)-t\|_\infty \to 0$, as $n\to\infty$,

(ii) $\sup_{t\in[-a,a]^d} |f_n(\lambda_n(t))-f(t)| \to 0$, as $n\to\infty$ for all $a>0$,

where $\|\cdot\|_\infty$ denotes the maximum norm in R^d.

Proof. See [ŁR86, Theorem 1] and [Fer15, p. 19]. □

Lemma 3.4 Let $d\in\mathbb{N}$ and define the map

$$\Phi : (D(\mathbb{R}), s_1) \to \left(D(\mathbb{R}^d), s_d\right), \quad f \mapsto \bar{f},$$

where $\bar{f}\in D(\mathbb{R}^d)$ with $(x_1,\dots,x_d)\mapsto \bar{f}(x_1,\dots,x_d) := f(x_1)$. Then Φ is continuous.

Proof. Let $f\in D(\mathbb{R})$ and $R=(R_1,\dots,R_d)\in\{<,\geq\}^d$. Since for each $t=(t_1,\dots,t_d)\in\mathbb{R}^d$ we have

$$\lim_{Q_R\ni s\to t} \bar{f}(s) = \lim_{Q_{R_1}\ni s_1\to t_1} f(s_1)\,,$$

and, in particular,

$$\lim_{Q_{(\geq,\dots,\geq)}\ni s\to t} \bar{f}(s) = \lim_{s_1\downarrow t_1} f(s_1) = f(t_1) = \bar{f}(t)\,,$$

$\bar{f}_{Q_R}$ exists and $\bar{f}_{Q_{(\geq,\dots,\geq)}}(t)=\bar{f}(t)$. Hence, $\bar{f}\in D(\mathbb{R}^d)$ and Φ is well-defined. In order to show continuity use the set of transformations Λ_d introduced in Lemma 3.3 and define the map $\varphi : \Lambda_1 \to \Lambda_d$ with $\lambda\mapsto\bar{\lambda}$, where $\bar{\lambda}\in\Lambda_d$ with $(t_1,\dots,t_d)\mapsto(\lambda(t_1),t_2,\dots,t_d)$. Then

$$\sup_{t\in\mathbb{R}^d} \left\|\bar{\lambda}(t)-t\right\|_\infty = \sup_{t_1\in\mathbb{R}} \|\lambda(t_1)-t_1\|_\infty$$

and for all $a>0$

$$\sup_{t\in[-a,a]^d} \left|\bar{f}(\bar{\lambda}(t))-\bar{f}(t)\right| = \sup_{t_1\in[-a,a]} |f(\lambda(t_1))-f(t_1)|\ .$$

By Lemma 3.3 the continuity of Φ follows immediately. □

Let $\mathscr{D}(\mathbb{R}^d)$ be the Borel σ-algebra generated by the Skorokhod topology s_d. A measurable map $X : (\Omega,\mathscr{A},\mathbb{P}) \to (D(\mathbb{R}^d),\mathscr{D}(\mathbb{R}^d))$ is called a stochastic process with trajectories in

$D(\mathbb{R}^d)$. Accordingly, the measurable map $X : (\Omega, \mathscr{A}, \mathbb{P}) \to (D(I^d), \mathscr{D}(I^d))$ is a stochastic process in $D(I^d)$ and we write $X|_{I^d}$ for the restriction. If $S \subseteq \mathbb{R}^d$ we will denote by $\pi_S(X)$ the projection of the process X onto its marginals, i.e. the family $\{X(s)\,; s \in S\}$.

We give a characterization of convergence in distribution of a sequence $(X_n)_{n\in\mathbb{N}}$ in $D(\mathbb{R}^d)$. The following proposition states that the convergence of a sequence of random variables in $D(\mathbb{R}^d)$ can be attributed to the convergence in $D(I^d)$.

Proposition 3.5 Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of stochastic processes in $D(\mathbb{R}^d)$. Then $X_n \xrightarrow{\mathscr{L}} X$, as $n \to \infty$, if and only if $X_n|_{[-a,a]} \xrightarrow{\mathscr{L}} X|_{[-a,a]}$ in $D([-a,a])$ for all $0 < a = (a_1, \ldots, a_d)$ with

$$a \in C_X := \left\{ t \in \mathbb{R}^d ; \pi_t \text{ continuous } \mathbb{P} \circ X^{-1}\text{-almost everywhere} \right\} .$$

Proof. The proof can be found in [ŁR86, Theorem 4]. □

In this thesis, we need an additional topological space to characterize trajectories of stochastic processes and their convergence.

Definition 3.6 (Skorokhod product space) Let $D_d := \prod_{i=1}^d D(\mathbb{R})$ be the d-fold cartesian product of $(D(\mathbb{R}), s_1)$ equipped with the product topology. Then we call D_d the *Skorokhod product space*. By $\mathscr{D}_d$ we denote the Borel σ-algebra generated by the product topology on D_d.

A common method to prove $X_n \xrightarrow{\mathscr{L}} X$ in $D(\mathbb{R})$ is initiated by Prokhorov's theorem and involves the procedure to prove first that the sequence $(X_n)_{n\in\mathbb{N}}$ is tight. Secondly, to show the convergence of the finite-dimensional marginal distributions (fidis), that means

$$\pi_S(X_n) \xrightarrow{\mathscr{L}} \pi_S(X)$$

for all finite $S \subseteq T$, where T is a dense subset of $\mathbb{R}$, see [JS03, Chapter VI, §3b, 3.20]. Now we formulate a corresponding weak convergence criterion in the Skorokhod product space D_d. The procedure is of a similar type as in [JS03, Chapter VI, §3b, 3.20] and is formulated in [FV09, Theorem 5.1].

Proposition 3.7 Let $(Z_n)_{n\in\mathbb{N}} = (X_n^{(1)}, \ldots, X_n^{(d)})_{n\in\mathbb{N}}$ be a sequence of stochastic processes in $(D_d, \mathscr{D}_d)$. If each of the sequences $X_n^{(1)}, \ldots, X_n^{(d)}$ is tight and there is a random variable $Z = (X^{(1)}, \ldots, X^{(d)})$ in $(D_d, \mathscr{D}_d)$ such that

$$\left(\pi_S\left(X_n^{(1)}\right), \ldots, \pi_S\left(X_n^{(d)}\right)\right) \xrightarrow{\mathscr{L}} \left(\pi_S\left(X^{(1)}\right), \ldots, \pi_S\left(X^{(d)}\right)\right) ,$$

as $n \to \infty$, for all finite $S \subseteq T_{X^{(1)}} \cap \ldots \cap T_{X^{(d)}}$, where

$$T_{X^{(i)}} := \{t \in \mathbb{R}; \pi_t \text{ continuous } \mathbb{P} \circ \left(X^{(i)}\right)^{-1}\text{-almost everywhere}\} ,$$

$1 \leq i \leq d$, then

$$Z_n \xrightarrow{\mathscr{L}} Z ,$$

as $n \to \infty$.

Proof. The statement is proved in [FV09, Theorem 5.1] for the case $d = 2$ and can be generalized without further ado. □

3.2 Convergence of Arginf- and Argsup-sets

Our first concern will be the sets of all infimizers and supremizers of a process in $D(\mathbb{R}^d)$.

Definition 3.8 Let X be a stochastic process with trajectories in $D(\mathbb{R}^d)$. We call the random set

$$\operatorname{Arginf}(X) := \left\{ t \in R^d;\ \min_{R \in \{<,\geq\}^d} X_{Q_R}(t) = \inf_{s \in \mathbb{R}^d} X(s) \right\}$$

the *set of all infimizers of* X and

$$\operatorname{Argsup}(X) := \left\{ t \in R^d;\ \max_{R \in \{<,\geq\}^d} X_{Q_R}(t) = \sup_{s \in \mathbb{R}^d} X(s) \right\}$$

the *set of all supremizers of* X. Accordingly we call the random sets

$$\operatorname{Argmin}(X) := \left\{ t \in R^d;\ X(t) = \inf_{s \in \mathbb{R}^d} X(s) \right\}$$

the *set of all minimizers of* X and

$$\operatorname{Argmax}(X) := \left\{ t \in R^d;\ X(t) = \sup_{s \in \mathbb{R}^d} X(s) \right\}$$

the *set of all maximizers of* X.

Lemma 3.9 Let X be a process in $D(\mathbb{R}^d)$ and $\Gamma = \operatorname{diag}(\alpha_1, \dots, \alpha_d)$ be a diagonal matrix with strictly positive entries. For all $\theta \in \mathbb{R}^d$ and $\gamma > 0$, then

$$Z(t) = \gamma \left\{ X(\theta + \Gamma^{-1} t) - X(\theta) \right\} ,$$

$t \in \mathbb{R}^d$, is a process in $D(\mathbb{R}^d)$ and

$$\Gamma \left(\hat{\theta} - \theta \right) \in \operatorname{Arginf}(Z)$$

for each $\hat{\theta} \in \operatorname{Arginf}(X)$.

Proof. Given the diagonal matrix Γ and $\theta \in R^d$ define a transformation $\lambda : \mathbb{R}^d \to \mathbb{R}^d$ with $\lambda(t) := \Gamma(t - \theta)$. Then λ is an element of the set Λ_d introduced in Lemma 3.3 and $Z(t) = \gamma \left\{ X(\lambda^{-1}(t)) - X(\theta) \right\}$, where λ^{-1} denotes the inverse of λ. Now the claim follows immediately from [Fer15, Lemma 2.2(i) and (iii)]. □

Remark 3.10 (i) As a direct consequence of the definition we have $\operatorname{Argmin}(X) \subseteq \operatorname{Arginf}(X)$ and $\operatorname{Argmax}(X) \subseteq \operatorname{Argsup}(X)$.

(ii) In the special case of stochastic processes with trajectories in $D(\mathbb{R})$, we obtain directly

$$\operatorname{Arginf}(X) = \left\{ t \in R;\ \min\{X(t-), X(t)\} = \inf_{s \in \mathbb{R}} X(s) \right\} ,$$
$$\operatorname{Argsup}(X) = \left\{ t \in R;\ \max\{X(t-), X(t)\} = \sup_{s \in \mathbb{R}} X(s) \right\} .$$

Moreover, due to $\sup(-A) = -\inf(A)$ for all $A \subseteq \mathbb{R}$, obviously $\operatorname{Arginf}(X) = \operatorname{Argsup}(-X)$.

In the following proposition we recall a sufficient condition for the convergence of infimizing points of processes in $D(\mathbb{R})$. Conditions to be adjusted to obtain an analogous result for supremizers are expressed in square brackets. Moreover, the result holds both for almost sure convergence as well as for convergence in probability, which is formulated in round brackets.

Proposition 3.11 Let X and X_n for $n \in \mathbb{N}$ be stochastic processes in $D(\mathbb{R})$ with

$$\sup_{t \in \mathbb{R}} |X_n(t) - X(t)| \longrightarrow 0 \quad \mathbb{P}\text{-a.s. (in probability) as } n \to \infty ,$$

and $\xi \in \mathrm{Arginf}(X)$ [or $\xi \in \mathrm{Argsup}(X)$] almost surely with

$$\inf_{t\in\mathbb{R}} X(t) < \inf\{X(t);\ |t-\xi| > \varepsilon\} \quad \left[\sup_{t\in\mathbb{R}} X(t) > \sup\{X(t);\ |t-\xi| > \varepsilon\}\right] \tag{3.1}$$

for all $\varepsilon > 0$ almost surely. For each sequence $(\xi_n)_{n\in\mathbb{N}}$ with

$$\xi_n \in \mathrm{Arginf}(X_n) \text{ almost surely for each } n \geq N$$
$$[\xi_n \in \mathrm{Argsup}(X_n) \text{ almost surely for each } n \geq N]$$

then

$$\xi_n \longrightarrow \xi \quad \mathbb{P}\text{-almost surely (in probability) as } n \to \infty\,.$$

Proof. This is a direct consequence of [Fer15, Theorem 3.3]. □

Condition (3.1) implies that $\mathrm{Arginf}(X) = \{\xi\}$ or $\mathrm{Argsup}(X) = \{\xi\}$, respectively. But in general, the limit process X has not a unique infimizing or supremizing point. For that reason we need a continuous mapping theorem for the Arginf- and Argsup-functional in a more general setting.

Proposition 3.12 Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of stochastic processes in $D(\mathbb{R}^d)$, and $(\xi_n)_{n\in\mathbb{N}}$ be a sequence with $\xi_n \in \mathrm{Arginf}(X_n)$ almost surely for each $n \geq N$. If $X_n \xrightarrow{\mathscr{L}} X$ in $D(\mathbb{R}^d)$, as $n \to \infty$, and $(\xi_n)_{n\in\mathbb{N}}$ is stochastically bounded, i.e.

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(|\xi_n| \geq k) = 0\,,$$

then

$$\limsup_{n\to\infty} \mathbb{P}(\xi_n \in F) \leq \mathbb{P}(\mathrm{Arginf}(X) \cap F \neq \emptyset) \tag{3.2}$$

for all closed sets $F \subseteq \mathbb{R}^d$.

Proof. See [Fer15, Theorem 3.11]. □

Remark 3.13 The expression on the right hand side of (3.2) is in general not a probability measure but a *Choquet-capacity functional* T_C of the so-called *random closed set* $C = \mathrm{Arginf}(X)$, this means that T_C is a set function defined on $\mathscr{B}(\mathbb{R}^d)$ and with $T_C(F) = \mathbb{P}(C \cap F \neq \emptyset)$. For further details confer with [Fer15, p. 35-37]. Due to the one-to-one correspondence of T_C and $\mathbb{P}\circ C^{-1}$ by Choquet's theorem (cf. [Fer15, Theorem 3.9]), one can characterize weak convergence of ξ_n in terms of the capacity functional T_C in the sense of [Fer15, Definition 3.1]. According to this definition we say $P_n := \mathbb{P}\circ\xi_n^{-1}$ *converges weakly to* T_C (or ξ_n *converges in distribution to* C, respectively) if $\limsup_{n\to\infty} P_n(F) \le T_C(F)$ for all closed sets $F \subseteq \mathbb{R}^d$. However, if $C = \{\xi\}$ almost surely then indeed $\mathbb{P}(C \cap \cdot \neq \emptyset) = \mathbb{P}(\xi \in \cdot)$ is a probability measure and we conclude from Portmanteau's theorem (cf. [Kal97, Theorem 3.25]), that $\xi_n \xrightarrow{\mathscr{L}} \xi$ in the classical sense. Thus, convergence to Choquet-capacity functionals is an extension of the notion of weak convergence.

The convergence in the sense of (3.2) is stable under continuous maps.

Proposition 3.14 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random variables in $\mathbb{R}^d$, and $C \subseteq \mathbb{R}^d$ be a random set which is compact almost surely with

$$\limsup_{n\to\infty} \mathbb{P}(\xi_n \in F) \le \mathbb{P}(C \cap F \neq \emptyset)$$

for all closed sets $F \subseteq \mathbb{R}^d$. If $h : \mathbb{R}^d \to \mathbb{R}^{d'}$, $d' \in \mathbb{N}$, is a measurable map such that $\mathbb{P}(C \cap D_h) = 0$, where D_h is the set of all discontinuity points of h, then $h(C)$ is compact almost surely and

$$\limsup_{n\to\infty} \mathbb{P}\big(h(\xi_n) \in F'\big) \le \mathbb{P}\big(h(C) \cap F' \neq \emptyset\big)$$

for all closed sets $F' \subseteq \mathbb{R}^{d'}$.

Proof. This is a special case of [Fer15, Propositon 3.4]. $\square$

4 Inequalities

In this chapter we introduce and investigate empirical processes appearing in the examination. By the usage of techniques and inequalities from martingale theory we establish supremal inequalities for them.

In Section 4.1 we will proof a type of Birnbaum and Marshall inequality for zero-mean martingales and, afterwards, a version for backwards martingales. These inequalities will be applied to the empirical processes E_n and $\bar{E}_n$ in Section 4.2 and 4.3. In Section 4.4 and 4.5 we utilize a Doob-Meyer-type decomposition for the empirical processes L_n and $\bar{L}_n$ and also derive supremal inequalities. Finally the results collected so far are utilized in Section 4.6 to show inequalities for the process W_n. If not stated otherwise, we work in the notation of Section 2.1.

Define the continuous-time processes $(E_n(t))_{t\in\mathbb{R}\cup\{-\infty\}}$ and $(\bar{E}_n(t))_{t\in\mathbb{R}\cup\{\infty\}}$ as

$$E_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\le t\}}\,(Y_i - m(X_i))\,, \quad t\in\mathbb{R}\,, \tag{4.1}$$

and

$$\bar{E}_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i> t\}}\,(Y_i - m(X_i))\,, \quad t\in\mathbb{R}\,, \tag{4.2}$$

where we set $E_n(-\infty)=0$ and $\bar{E}_n(\infty)=0$. Furthermore, define the continuous-time processes $(L_n(t))_{t\in\mathbb{R}\cup\{-\infty\}}$ and $(\bar{L}_n(t))_{t\in\mathbb{R}\cup\{\infty\}}$ as

$$L_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i\le t\}}m(X_i)\,, \quad t\in\mathbb{R}\,, \tag{4.3}$$

and

$$\bar{L}_n(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i> t\}}m(X_i)\,, \quad t\in\mathbb{R}\,, \tag{4.4}$$

where $L_n(-\infty) = 0$ and $\bar{L}_n(\infty) = 0$. We will also introduce the continuous-time process $(W_n(t))_{t\in\mathbb{R}}$, where

$$W_n(t) := \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{1}_{X_i\le t}\left(Y_i - \frac{\alpha+\beta}{2}\right) - \mathbb{E}\left(\mathbb{1}_{X_i\le t}\left(Y_i - \frac{\alpha+\beta}{2}\right)\right)\right]. \tag{4.5}$$

In the sequel, the tuple $(X_{i:n}, Y_{[i:n]})$ denotes the i—th order statistic $X_{i:n}$ and its corresponding concomitant $Y_{[i:n]}$, this means that $X_{1:n} \le \ldots \le X_{n:n}$, and $Y_{[i:n]}$ is the random variable associated with $X_{i:n}$. Moreover, we set $\underline{X}_n := (X_{1:n}, \ldots, X_{n:n})$ and $\underline{Y}_n := (Y_{[1:n]}, \ldots, Y_{[n:n]})$. We adopt the convention that $\sup_{t\in\emptyset} f(t) = -\infty$.

Remark 4.1 (Generic constant) In order to prevent excessive redefinition of constants in the following estimations, we will use the concept of a generic positive constant. We call $D > 0$ a generic positive constant if it represents a strictly positive term which may change from line to line. It does not depend on the involved parameters and, therefore, must not be known exactly.

4.1 Birnbaum and Marshall inequality for zero-mean martingales and backwards martingales

The following inequality represents a generalization of Inequality 4 in [SW86, A.10.].

Lemma 4.2 Let $(S_t, \mathscr{F}_t)_{\gamma\le t\le\delta}$, be a zero-mean, continuous-time and square-integrable martingale with càdlàg sample paths. If $q : [\gamma, \delta] \to (0, \infty)$ is a non-decreasing and right-continuous function, then for all $\lambda > 0$

$$\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|S_t|}{q(t)} \ge \lambda\right) \le \lambda^{-2}\left\{q(\gamma)^{-2}\mu(\gamma) + \int_{(\gamma,\delta]} q(t)^{-2}\mu(\mathrm{d}t)\right\}, \tag{4.6}$$

where

$$\mu(t) := \begin{cases} \mathbb{E}S_\gamma^2, & t < \gamma \\ \mathbb{E}S_t^2, & t \in [\gamma, \delta] \\ \mathbb{E}S_\delta^2, & t > \delta \end{cases}.$$

Under the same assumptions but $(S_t, \mathscr{F}_t)$ being left continuous with right limits (càglàd) and q being left-continuous then

$$\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|S_t|}{q(t)}\ge\lambda\right)\le\lambda^{-2}\left\{q(\gamma)^{-2}\mu(\gamma)+\int\limits_{[\gamma,\delta)}q(t+)^{-2}\mu(\,\mathrm{d}t)\right\}\,. \tag{4.7}$$

Remark 4.3 Comparing the inequality in [SW86, A.10.] with (4.7) we find a small discrepancy in the term in the integral. Indeed the càdlàg and càglàd cases can be dealt separately to sharpen the inequality. But nevertheless in section 4.5 we will need this more strict version of inequality.

Proof. The proof for $\gamma = 0$ and $S_\gamma = 0$ can be found in Inequality 4 in [SW86, A.10.]. We follow the line of argument there. First consider the right-continuous case. For each $n \in \mathbb{N}$ let $\{v_k := \gamma + (\delta-\gamma)(k-1)2^{-n}\,; 1 \le k \le 2^n+1\}$ be a partition of $[\gamma,\delta]$, which refines $[\gamma,\delta]$ into 2^n subintervals. Define a discrete time process $(B_k, \mathscr{G}_k)_{0\le k\le 2^n+1}$ with $B_0 := 0, B_k := S_{v_k}$ and $\mathscr{G}_0 := \{\Omega, \emptyset\}$, $\mathscr{G}_k := \mathscr{F}_{v_k}$. Observe that $(B_k, \mathscr{G}_k)_{0\le k\le 2^n+1}$ is a zero-mean martingale. Fix $\lambda > 0$. By the right-continuity of $S_t/q(t)$ we get

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|S_t|}{q(t)}>\lambda\right)\\
&=\mathbb{P}\left(\sup_{n\in\mathbb{N}}\max_{1\le k\le 2^n+1}\frac{|B_k|}{q(v_k)}>\lambda\right)\\
&=\mathbb{P}\left(\bigcup_{n\in\mathbb{N}}\left\{\max_{1\le k\le 2^n+1}\frac{|B_k|}{q(v_k)}>\lambda\right\}\right)\\
&\overset{(a)}{=}\lim_{n\to\infty}\mathbb{P}\left(\left\{\max_{1\le k\le 2^n+1}\frac{|B_k|}{q(v_k)}>\lambda\right\}\right)\\
&\overset{(b)}{\le}\lim_{n\to\infty}\lambda^{-2}\left\{q(v_1)^{-2}\mathrm{Var}(B_1)+\sum_{k=2}^{2^n+1}q(v_k)^{-2}\mathrm{Var}(B_k-B_{k-1})\right\}\\
&\overset{(c)}{=}\lambda^{-2}\left\{q(\gamma)^{-2}\mathbb{E}S_\gamma^2+\lim_{n\to\infty}\sum_{k=2}^{2^n+1}q(v_k)^{-2}\left(\mathbb{E}S_{v_k}^2-\mathbb{E}S_{v_{k-1}}^2\right)\right\}\\
&\overset{(d)}{=}\lambda^{-2}\left\{q(\gamma)^{-2}\mu(\gamma)+\int\limits_{(\gamma,\delta]}q(t)^{-2}\mu(\,\mathrm{d}t)\right\}\,,
\end{aligned}$$

where in (a) we use the continuity from below of $\mathbb{P}$ and the fact that $\left\{\max_{1\le k\le 2^n+1}\frac{|B_k|}{q(v_k)}>\lambda\right\}$ is non-decreasing in n. In (b) we use Inequality A.3 and in (c) we make use of the zero-mean martingale property. To see the convergence in (d) observe that $\mathbb{E}S_t^2$ is non-negative, non-decreasing and right continuous, and thus, $\mu(t)$ measure-defining. This means there corresponds exactly one measure, which we also denote by μ, satisfying $\mu((a,b]) = \mathbb{E}S_b^2 - \mathbb{E}S_a^2$ for all $(a,b] \subseteq [\gamma,\delta]$, see [Bil95, theorem 12.4]. For each $n \in \mathbb{N}$ identify

$$\sum_{k=2}^{2^n+1} q(v_k)^{-2}\left(\mathbb{E}S_{v_k}^2 - \mathbb{E}S_{v_{k-1}}^2\right) = \int h_n(t)\mu(\,\mathrm{d}t)\,,$$

where

$$h_n(t) := \sum_{k=2}^{2^n+1} q(v_k)^{-2}\mathbb{1}_{(v_{k-1},v_k]}(t)\,.$$

Since $(h_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of positive and measurable functions with $\lim_{n\to\infty} h_n(t) = q^{-2}(t+) = q^{-2}(t)$ for all $t \in (\gamma,\delta]$ the convergence follows from an application of the monotone convergence theorem. The left continuous case only differs in (d). Since $\mathbb{E}S_t^2$ is non-decreasing and left continuous, there exists exactly one measure, denoted by μ, satisfying $\mu([a,b)) = \mathbb{E}S_b^2 - \mathbb{E}S_a^2$ for all $[a,b) \subseteq [\gamma,\delta]$. Now, for each $n \in \mathbb{N}$ identify

$$\sum_{k=2}^{2^n+1} q(v_k)^{-2}\left(\mathbb{E}S_{v_k}^2 - \mathbb{E}S_{v_{k-1}}^2\right) = \int h_n(t)\mu(\,\mathrm{d}t)\,,$$

where

$$h_n(t) := \sum_{k=2}^{2^n+1} q(v_k)^{-2}\mathbb{1}_{[v_{k-1},v_k)}(t)\,.$$

Since $(h_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of positive and measurable functions with $\lim_{n\to\infty} h_n(t) = q^{-2}(t+)$ for all $t \in [\gamma,\delta)$ the convergence can be seen by applying the monotone convergence theorem. □

We will formulate an appropriate inequality for the case, where S_t is a backwards martingale.

Lemma 4.4 Let $(S_t,\mathscr{F}_t)_{\gamma\le t\le\delta}$, be a zero-mean and square-integrable backwards martingale with càdlàg sample paths. Let $q : [\gamma,\delta] \to (0,\infty)$ be a non-increasing right continuous function. Then for all $\lambda > 0$

$$\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|S_t|}{q(t)} \ge \lambda \right) \le \lambda^{-2} \left\{ q(\delta)^{-2} \mu(\delta) + \int_{(\gamma,\delta]} q(t-)^{-2} (-\mu(\,\mathrm{d}t)) \right\} , \tag{4.8}$$

where

$$\mu(t) := \begin{cases} \mathbb{E}S_\gamma^2\,, & t < \gamma \\ \mathbb{E}S_t^2\,, & t \in [\gamma, \delta] \\ \mathbb{E}S_\delta^2\,, & t > \delta \end{cases} .$$

Under the same assumptions but $(S_t, \mathscr{F}_t)$ being càglàd and q being left continuous then

$$\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|S_t|}{q(t)} \ge \lambda \right) \le \lambda^{-2} \left\{ q(\delta)^{-2} \mu(\delta) + \int_{[\gamma,\delta)} q(t)^{-2} (-\mu(\,\mathrm{d}t)) \right\} .$$

Proof. First, consider the right-continuous case. Set $\tilde{S}_t := S_{\delta - t + \gamma}$, $\tilde{\mathscr{F}}_t := \mathscr{F}_{\delta - t + \gamma}$ and $\tilde{q}(t) := q(\delta - t + \gamma)$. Note that $(\tilde{S}_t, \tilde{\mathscr{F}}_t)$ is a zero-mean and square integrable martingale with càglàd sample paths and $\tilde{q}(t)$ is a non-decreasing and left continuous function. Applying Lemma 4.2 yields

$$\begin{aligned} &\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|S_t|}{q(t)} > \lambda \right) \\ &= \mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|\tilde{S}_t|}{\tilde{q}(t)} > \lambda \right) \\ &\le \lambda^{-2} \left\{ \tilde{q}(\gamma)^{-2} \tilde{\mu}(\gamma) + \int_{[\gamma,\delta)} \tilde{q}(t+)^{-2} \tilde{\mu}(\,\mathrm{d}t) \right\} , \end{aligned}$$

where $\mathbb{E}\tilde{S}_t^2$ is the measure defining function for the measure $\tilde{\mu}$ with $\tilde{\mu}([a,b)) = \mathbb{E}\tilde{S}_b^2 - \mathbb{E}\tilde{S}_a^2$. If we define $\phi(t) := \delta - t + \gamma$ we obtain

$$\tilde{\mu}([a,b)) = \mathbb{E}\tilde{S}_b^2 - \mathbb{E}\tilde{S}_a^2 = -(\mathbb{E}S_{\phi(a)}^2 - \mathbb{E}S_{\phi(b)}^2) = -\mu((\phi(b), \phi(a)]) .$$

We continue

$$
\begin{aligned}
&\lambda^{-2}\left\{\tilde{q}(\gamma)^{-2}\tilde{\mu}(\gamma)+\int\limits_{[\gamma,\delta)}\tilde{q}(t+)^{-2}\tilde{\mu}(\mathrm{d}t)\right\}\\
&=\lambda^{-2}\left\{\tilde{q}(\gamma)^{-2}\tilde{\mu}(\gamma)+\int \mathbb{1}_{[\gamma,\delta)}(t)\left(\lim_{s\downarrow t}\tilde{q}(s)\right)^{-2}\tilde{\mu}(\mathrm{d}t)\right\}\\
&=\lambda^{-2}\left\{q(\delta)^{-2}\mu(\delta)+\int \mathbb{1}_{(\gamma,\delta]}(\phi(t))\left(\lim_{\phi(s)\uparrow\phi(t)}q(\phi(s))\right)^{-2}(-\mu(\mathrm{d}\phi(t)))\right\}\\
&=\lambda^{-2}\left\{q(\delta)^{-2}\mu(\delta)+\int\limits_{(\gamma,\delta]}q(u-)^{-2}(-\mu(\mathrm{d}u))\right\}\ .
\end{aligned}
$$

The proof for the left continuous case is quite similar. We only have to replace càglàd by càdlàg, left continuous by right-continuous and intervals of the form $[a,b)$ by $(a,b]$ and use Inequality (4.6) from Lemma 4.2 instead of Inequality (4.7). □

4.2 Results for E_n

Lemma 4.5 Suppose that $\gamma\in\mathbb{R}$ and $q:[\gamma,\infty)\to(0,\infty)$ is a right-continuous and monotonically increasing function. Let $u\in\{-\infty\}\cup(-\infty,\gamma]$ and $\delta>\gamma$. If F is continuous on $(\gamma,\delta]$ then the inequality

$$
\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|E_n(t)-E_n(u)|}{q(t)}\ge\varepsilon\right)\le n^{-1}\varepsilon^{-2}\left[2q(\gamma)^{-2}\int\limits_{(u,\gamma]}V(x)\,F(\mathrm{d}x)+\int\limits_{(\gamma,\delta]}q(x)^{-2}V(x)F(\mathrm{d}x)\right]
$$

holds for all $\varepsilon>0$.

The argumentation in the following proof was applied to different special cases as in [FK09, p. 103-104], [Til07, p. 59-70] and [Fer09, Ch. 8].

Proof. Define the set $\Omega_n\subseteq\Omega$, where $\Omega_n:=\bigcup_{k=1}^{n}\{X_k\in(\gamma,\delta]\}$. Furthermore, set

$$
\begin{aligned}
T_n&:=\left\{\underline{x}_n=(x_1,\ldots,x_n)\in\mathbb{R}^n;\ x_1\le\ldots\le x_n\right\}\\
R_n&:=\left\{\underline{x}_n\in T_n;\ \exists k\in\{1,\ldots,n\}\ \text{with}\ x_k\in(\gamma,\delta]\right\}\ ,
\end{aligned}
$$

and for all $\underline{x}_n \in R_n$ define

$$
\begin{aligned}
a(\underline{x}_n) &:= \min\{1 \le l \le n \,;\, x_l \in (\gamma, \delta]\} \\
b(\underline{x}_n) &:= \max\{1 \le l \le n \,;\, x_l \in (\gamma, \delta]\} \\
c(\underline{x}_n) &:= \sum_{i=1}^{n} \mathbb{1}_{\{x_i \le u\}} \,.
\end{aligned}
$$

For the vector of order statistics $\underline{X}_n$ then $\Omega_n = \underline{X}_n^{-1}(R_n) \subseteq \underline{X}_n^{-1}(T_n)$ and, hence, $\Omega_n \in \underline{X}_n^{-1}(\mathscr{B}(T_n)) = \sigma(\underline{X}_n)$ and $\mathbb{1}_{\Omega_n}$ is $\sigma(\underline{X}_n)$-measurable. Additionally, on the set Ω_n the random variables $A := a(\underline{X}_n)$ and $B := b(\underline{X}_n)$ are well-defined. By commutativity of summation and the fact that $X_{i:n} \le t$ if and only if $i \le nF_n(t)$ (cf. [Geo13, Remark 8.16]), we can write

$$
\begin{aligned}
E_n(t) &\overset{(4.1)}{=} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}} (Y_i - m(X_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_{i:n} \le t} \left(Y_{[i:n]} - m(X_{i:n})\right) \\
&= \frac{1}{n}\sum_{i=1}^{nF_n(t)} \left(Y_{[i:n]} - m(X_{i:n})\right) .
\end{aligned}
$$

If we set

$$
\Delta_n(t) := E_n(t) - E_n(u) = \frac{1}{n}\sum_{i=nF_n(u)+1}^{nF_n(t)} \left(Y_{[i:n]} - m(X_{i:n})\right) ,
$$

we observe that Δ_n is a càdlàg process which is constant on each interval $[X_{i:n}, X_{i+1:n})$ and with jumps in each $X_{i:n}$. Since $t \mapsto 1/q(t)$ is monotonically decreasing, the process $\left\{|\Delta_n(t)|\, q(t)^{-1}; \gamma \le t \le \delta\right\}$ is monotonically decreasing within $[X_{i:n}, X_{i+1:n})$ and with jumps occurring at most in $X_{i:n}$ or in the discontinuity points of q. In the latter case the jump size is negative. This implies that all suprema of this process within $[X_{i:n}, X_{i+1:n})$ appear at the left boundary $X_{i:n}$. Hence, for all $\omega \in \Omega_n$ then

$$
\sup_{\gamma \le t \le \delta} \frac{|\Delta_n(t)|}{q(t)} = \max\left\{ \frac{|\Delta_n(\gamma)|}{q(\gamma)},\ \max_{A \le k \le B} \frac{|\Delta_n(X_{k:n})|}{q(X_{k:n})} \right\} .
$$

In contrast, for all $\omega \notin \Omega_n$ the process Δ_n is constant on $(\gamma, \delta]$ and thus

$$\sup_{\gamma \le t \le \delta} \frac{|\Delta_n(t)|}{q(t)} = \frac{|\Delta_n(\gamma)|}{q(\gamma)} .$$

Hence,

$$\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|\Delta_n(t)|}{q(t)} \ge \varepsilon \right) \le \mathbb{P}\left(\frac{|\Delta_n(\gamma)|}{q(\gamma)} \ge \varepsilon \right) + \mathbb{P}\left(\Omega_n \cap \left\{ \max_{A \le k \le B} \frac{|\Delta_n(X_{k:n})|}{q(X_{k:n})} \ge \varepsilon \right\} \right) . \tag{4.9}$$

For the first summand on the right hand side of (4.9) we use Chebychev's inequality and the fact that the summands of $\Delta_n(\gamma)$ are i. i. d. and zero-mean to obtain

$$\begin{aligned}
&\mathbb{P}\left\{ \frac{|\Delta_n(\gamma)|}{q(\gamma)} \ge \varepsilon \right\} \\
&= \mathbb{P}\left\{ \left| \sum_{i=1}^{n} \mathbb{1}_{X_i \in (u,\gamma]} (Y_i - m(X_i)) \right| \ge \varepsilon q(\gamma) n \right\} \\
&\le (\varepsilon q(\gamma) n)^{-2} \operatorname{Var}\left(\sum_{i=1}^{n} \mathbb{1}_{X_i \in (u,\gamma]} (Y_i - m(X_i)) \right) \\
&= (\varepsilon q(\gamma))^{-2} n^{-1} \mathbb{E}\left(\mathbb{1}_{X \in (u,\gamma]} (Y - m(X))^2 \right) \\
&= (\varepsilon q(\gamma))^{-2} n^{-1} \int_{(u,\gamma]} V(x) F(\mathrm{d}x) . \qquad (4.10)
\end{aligned}$$

Now we study the second summand on the right hand side of (4.9) and prove a Hájek-Rényi like inequality for order statistics. Since F is continuous on the interval $(\gamma, \delta]$, with probability one there are no ties in $(\gamma, \delta]$. Thus, for all $A \le k \le B$ then $nF_n(X_{k:n}) = k$ (cf. [Geo13, Remark 8.16]) and

$$\Delta_n(X_{k:n}) = \frac{1}{n} \sum_{i=nF_n(u)+1}^{nF_n(X_{k:n})} \left(Y_{[i:n]} - m(X_{i:n}) \right) = \frac{1}{n} \sum_{i=C+1}^{k} \left(Y_{[i:n]} - m(X_{i:n}) \right) ,$$

where $C := c(\underline{X}_n)$. Now define the mapping $H : T_n \times \mathbb{R}^n \to \mathbb{R}$ with $H(\underline{x}_n, \underline{y}_n) = \mathbb{1}_M$, where

$$M := \left\{ \max_{a(\underline{x}_n) \le k \le b(\underline{x}_n)} \frac{\left| \sum_{i=c(\underline{x}_n)+1}^{k} y_i - m(x_i) \right|}{q(x_k)} \ge n\varepsilon \right\} .$$

From the above it follows that

$$\begin{aligned}
&\mathbb{P}\left(\Omega_n \cap \left\{\max_{A\le k\le B} \frac{|\Delta_n(X_{k:n})|}{q(X_{k:n})} \ge \varepsilon\right\}\right) \\
&= \mathbb{E}\left(\mathbb{1}_{\Omega_n} H\left(\underline{X}_n, \underline{Y}_n\right)\right) \\
&= \int_{\Omega_n} \mathbb{E}\left(H\left(\underline{X}_n, \underline{Y}_n\right) | \underline{X}_n\right) \, \mathrm{d}\mathbb{P} \\
&= \int_{R_n} \mathbb{E}\left(H\left(\underline{X}_n, \underline{Y}_n\right) | \underline{X}_n = \underline{x}_n\right) \left(\mathbb{P} \circ \underline{X}_n^{-1}\right)(\mathrm{d}\underline{x}_n)\,.
\end{aligned}$$

By the lemma of Stute and Wang (cf. A.1), $\mathbb{P}_{\underline{Y}_n|\underline{X}_n}(\underline{x}_n, \mathrm{d}\underline{y}_n) = \bigotimes_{i=1}^n \mathbb{P}_{Y|X}(x_i, \mathrm{d}y_i)$, and by a characterization of conditional expectations, see [Als98, Proposition 53.6], for $\mathbb{P} \circ \underline{X}_n^{-1}$-almost all $\underline{x}_n \in R_n$ one has

$$\begin{aligned}
&\mathbb{E}\left(H(\underline{X}_n, \underline{Y}_n) | \underline{X}_n = \underline{x}_n\right) \\
&= \int H(\underline{x}_n, \underline{y}_n) \mathbb{P}_{\underline{Y}_n|\underline{X}_n}\left(\underline{x}_n, \mathrm{d}\,\underline{y}_n\right) \\
&= \int H(\underline{x}_n, \underline{y}_n) \bigotimes_{i=1}^n \mathbb{P}_{Y|X}\,(x_i, \mathrm{d}y_i)\,.
\end{aligned}$$

For $\underline{x}_n \in R_n$ let $W_1, \ldots, W_n$ be some independent random variables with $\mathbb{P} \circ (W_i)^{-1} = \mathbb{P}_{Y|X}(x_i, \cdot)$, $1 \le i \le n$. For brevity set $a := a(\underline{x}_n)$, $b := b(\underline{x}_n)$ and $c := c(\underline{x}_n)$, then

$$\begin{aligned}
&\int H(\underline{x}_n, \underline{y}_n) \bigotimes_{i=1}^n \mathbb{P}_{Y|X}\,(x_i, \mathrm{d}y_i) \\
&= \int H(\underline{x}_n, \underline{y}_n) \mathbb{P} \circ (W_1, \ldots, W_n)^{-1}\left(\mathrm{d}\underline{y}_n\right) \\
&= \mathbb{P}\left(\max_{a\le k\le b} \frac{\left|\sum_{i=c+1}^{k} (W_i - m(x_i))\right|}{q(x_k)} \ge n\varepsilon\right).
\end{aligned}$$

Since $\mathbb{E}(W_i) = \int y\, \mathbb{P}_{Y|X}(x_i, \mathrm{d}y) = \mathbb{E}(Y|X = x_i) = m(x_i)$, the summands are zero-mean with $\mathrm{Var}(W_i) = \int (y - m(x_i))^2 \mathbb{P}_{Y|X}(x_i, \mathrm{d}y) = V(x_i)$. Furthermore, $q(x_a) \le \ldots \le q(x_b)$

and after substituting $l = k - c$ we can use inequality A.2 of Hájek and Rényi and get

$$\begin{aligned}
&\mathbb{P}\left(\max_{a \le k \le b} \frac{\left| \sum_{i=c+1}^{k} (W_i - m(x_i)) \right|}{q(x_k)} \ge n\varepsilon \right) \\
&= \mathbb{P}\left(\max_{a-c \le l \le b-c} \frac{\left| \sum_{i=1}^{l} (W_{i+c} - m(x_{i+c})) \right|}{q(x_{l+c})} \ge n\varepsilon \right) \\
&\le (n\varepsilon)^{-2} \left\{ q(x_a)^{-2} \sum_{i=c+1}^{a} V(x_i) + \sum_{i=a+1}^{b} q(x_i)^{-2} V(x_i) \right\} \\
&= (n\varepsilon)^{-2} \left\{ q(x_a)^{-2} \sum_{i=c+1}^{a-1} V(x_i) + \sum_{i=a}^{b} q(x_i)^{-2} V(x_i) \right\} .
\end{aligned}$$

Observe that $X_{C:n} \le u < X_{C+1:n}$ and for each $\omega \in \Omega_n$,

$$X_{A-1:n} \le \gamma < X_{A:n} \le \ldots \le X_{B:n} \le \delta < X_{B+1:n} .$$

Therefore, we have $\mathbb{1}_{\Omega_n} q(X_{A:n})^{-2} \le q(\gamma)^{-2}$ and

$$\begin{aligned}
\mathbb{1}_{\Omega_n} \sum_{i=C+1}^{A-1} V(X_{i:n}) &= \sum_{i=1}^{n} \mathbb{1}_{X_{i:n} \in (u,\gamma]} V(X_{i:n}) = \sum_{i=1}^{n} \mathbb{1}_{X_i \in (u,\gamma]} V(X_i) \\
\mathbb{1}_{\Omega_n} \sum_{i=A}^{B} q(X_{i:n})^{-2} V(X_{i:n}) &= \sum_{i=1}^{n} \mathbb{1}_{X_{i:n} \in (\gamma,\delta]} q(X_{i:n})^{-2} V(X_{i:n}) = \sum_{i=1}^{n} \mathbb{1}_{X_i \in (\gamma,\delta]} q(X_i)^{-2} V(X_i) .
\end{aligned}$$

Finally,

$$\begin{aligned}
&\mathbb{P}\left(\Omega_n \cap \left\{ \max_{A \le k \le B} \frac{|\Delta_n(X_{k:n})|}{q(X_{k:n})} \ge \varepsilon \right\} \right) \\
&\le \int_{R_n} (n\varepsilon)^{-2} \left\{ q(x_a)^{-2} \sum_{i=c+1}^{a-1} V(x_i) + \sum_{i=a}^{b} q(x_i)^{-2} V(x_i) \right\} \mathbb{P} \circ \underline{X}_n^{-1}(\mathrm{d}\underline{x}_n) \\
&= (n\varepsilon)^{-2} \left\{ \mathbb{E}\left(\mathbb{1}_{\Omega_n} q(X_{A:n})^{-2} \sum_{i=C+1}^{A-1} V(X_{i:n}) \right) + \mathbb{E}\left(\mathbb{1}_{\Omega_n} \sum_{i=A}^{B} q(X_{i:n})^{-2} V(X_{i:n}) \right) \right\}
\end{aligned}$$

$$\leq (n\varepsilon)^{-2}\left\{\mathbb{E}\left(q(\gamma)^{-2}\sum_{i=1}^{n}\mathbb{1}_{X_i\in(u,\gamma]}V(X_i)\right)+\mathbb{E}\left(\sum_{i=1}^{n}\mathbb{1}_{X_i\in(\gamma,\delta]}q(X_i)^{-2}V(X_i)\right)\right\}$$
$$= n^{-1}\varepsilon^{-2}\left\{q(\gamma)^{-2}\int\limits_{(u,\gamma]}V(x)F(\,\mathrm{d}x)+\int\limits_{(\gamma,\delta]}q(x)^{-2}V(x)F(\,\mathrm{d}x)\right\},$$

and, together with (4.9) and (4.10), the proof is complete. □

Corollary 4.6 If $\nu > 0$ and if $r > 0$ such that **(A2)-(A3)** holds for $U_r(\tau)$, then there exists a constant $C > 0$ such that

$$\limsup_{n\to\infty}\mathbb{P}\left(\sup_{d\leq t\leq n^{\nu}r}\frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^{\frac{\nu+1}{2\nu}}}\geq\varepsilon\right)\leq\varepsilon^{-2}d^{-\frac{1}{\nu}}C$$

for all $\varepsilon > 0$ and $d > 0$. Moreover,

$$\limsup_{n\to\infty}\mathbb{P}\left(\sup_{d\leq t\leq n^{\nu}r}\frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^{\lambda}}\geq\varepsilon\right)=0$$

for all $\varepsilon > 0$, $d > 0$ and $1/2 \leq \lambda < \nu+1/2\nu$.

Proof. Fix $\varepsilon > 0$ and $d > 0$. For a given $\nu > 0$, let $\lambda \in \mathbb{R}$ be such that $1/2 \leq \lambda \leq \nu+1/2\nu$. Replace $s = \tau + n^{-\nu}t$ and note that

$$\mathbb{P}\left(\sup_{d\leq t\leq n^{\nu}r}\frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^{\lambda}}\geq\varepsilon\right)=\mathbb{P}\left(\sup_{\tau+n^{-\nu}d\leq s\leq\tau+r}\frac{|E_n(s)-E_n(\tau)|}{(s-\tau)^{\lambda}}\geq\varepsilon\right).$$

As required in Lemma 4.5, the function $q(s) = (s-\tau)^{\lambda}$ is positive, continuous, and monotonically increasing on $[\tau + n^{-\nu}d, \infty)$. By Lemma 4.5 and under assumptions **(A2)** and **(A3)** then for sufficiently large $n \in \mathbb{N}$

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\tau+n^{-\nu}d \leq s \leq \tau+r} \frac{|E_n(s)-E_n(\tau)|}{(s-\tau)^{\lambda}} \geq \varepsilon \right) \\
&\leq n^{-1}\varepsilon^{-2} \left[2(n^{-\nu}d)^{-2\lambda} \int\limits_{(\tau,\tau+n^{-\nu}d]} V(x)F(\mathrm{d}x) + \int\limits_{(\tau+n^{-\nu}d,\tau+r]} (x-\tau)^{-2\lambda}V(x)F(\mathrm{d}x) \right] \\
&\leq \varepsilon^{-2} \left[2n^{2\lambda\nu-\nu-1}d^{-2\lambda+1} \frac{\int_{(\tau,\tau+n^{-\nu}d]} V(x)F(\mathrm{d}x)}{F(\tau+n^{-\nu}d)-F(\tau)} \frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d} \right. \\
&\qquad\qquad \left. + n^{-1}\|V\|_{U_r(\tau)} \bar{L}^{2\lambda} \int\limits_{(\tau+n^{-\nu}d,\tau+r]} (F(x)-F(\tau))^{-2\lambda}F(\mathrm{d}x) \right] .
\end{aligned}
$$

We consider the second integral on the right-hand side of this inequality and assume $1/2 < \lambda \leq \nu+1/2\nu$. The case $\lambda = 1/2$ is treated separately. By assumption **(A2)**, F is strictly increasing and continuous on $[\tau, \tau+r]$. Hence, the integrand is also Riemann-Stieltjes integrable (cf. [SP12, A.36(e)]). In this case both constructions, Riemann-Stieltjes and Lebesgue integral, coincide (cf. [SP12, A.36(i)]). To distinguish these constructions different notations for the integration boundaries are used. Now we can apply the change-of-variable theorem for Riemann-Stieltjes integrals [Rud76, Theorem 6.19]. For this purpose set $g(x) = \frac{1}{-2\lambda+1}(x-F(\tau))^{-2\lambda+1}$ and note that $g'(x) = (x-F(\tau))^{-2\lambda}$, then

$$
\begin{aligned}
&\int\limits_{(\tau+n^{-\nu}d,\tau+r]} (F(x)-F(\tau))^{-2\lambda}F(\mathrm{d}x) \\
&= \int\limits_{\tau+n^{-\nu}d}^{\tau+r} g'(F(x))F(\mathrm{d}x) \\
&= \int\limits_{F(\tau+n^{-\nu}d)}^{F(\tau+r)} g'(x)\,\mathrm{d}x \\
&= g(F(\tau+r)) - g(F(\tau+n^{-\nu}d)) \\
&= \frac{1}{2\lambda-1}\left[\left(F(\tau+n^{-\nu}d)-F(\tau)\right)^{-2\lambda+1} - (F(\tau+r)-F(\tau))^{-2\lambda+1} \right] . \qquad (4.11)
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{d\le t\le n^\nu r}\frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^\lambda}\ge\varepsilon\right)\\
&\le\varepsilon^{-2}\left[2n^{2\lambda\nu-\nu-1}d^{-2\lambda+1}\frac{\int\limits_{(\tau,\tau+n^{-\nu}d]}V(x)F(\,\mathrm{d}x)}{F(\tau+n^{-\nu}d)-F(\tau)}\,\frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d}\right.\\
&\qquad+\|V\|_{U_r(\tau)}\,\bar{L}^{2\lambda}\frac{1}{2\lambda-1}\left[n^{2\lambda\nu-\nu-1}d^{-2\lambda+1}\left(\frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d}\right)^{-2\lambda+1}\right.\\
&\qquad\qquad\left.\left.-n^{-1}\left(F(\tau+r)-F(\tau)\right)^{-2\lambda+1}\right]\right]\,. \qquad (4.12)
\end{aligned}
$$

If we have $\lambda = {}^{\nu+1}/_{2\nu}$, then use A.5 to deduce

$$
\begin{aligned}
&\limsup_{n\to\infty}\mathbb{P}\left(\sup_{d\le t\le n^\nu r}\frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^{\frac{\nu+1}{2\nu}}}\ge\varepsilon\right)\\
&\le\varepsilon^{-2}d^{-\frac{1}{\nu}}\left[2V(\tau+)F'(\tau+)+\|V\|_{U_r(\tau)}\,\bar{L}^{\frac{\nu+1}{\nu}}\,\nu\left(F'(\tau+)\right)^{-\frac{1}{\nu}}\right],
\end{aligned}
$$

where the expressions in the squared brackets is chosen to be the constant C. Otherwise, if we have ${}^1/_2 < \lambda < {}^{\nu+1}/_{2\nu}$, then

$$
\limsup_{n\to\infty}\mathbb{P}\left(\sup_{d\le t\le n^\nu r}\frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^\lambda}\ge\varepsilon\right)=0\,.
$$

In the case $\lambda = {}^1/_2$ set $g(x)=\ln(x-F(\tau))$ and, according to this, $g'(x)=(x-F(\tau))^{-1}$. With similar arguments used to derive (4.11) we find

$$
\int\limits_{(\tau+n^{-\nu}d,\tau+r]}(F(x)-F(\tau))^{-1}F(\,\mathrm{d}x)=\ln\left(F(\tau+r)-F(\tau)\right)-\ln\left(F(\tau+n^{-\nu}d)-F(\tau)\right)\,.
$$

Hence,

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le t\le n^\nu r} \frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^{\frac{1}{2}}} \ge \varepsilon \right) \\
&\le \limsup_{n\to\infty} \varepsilon^{-2} \left[2n^{-1} \frac{\int\limits_{(\tau,\tau+n^{-\nu}d]} V(x)F(\mathrm{d}x)}{F(\tau+n^{-\nu}d)-F(\tau)} \frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d} \right. \\
&\qquad \left. + \|V\|_{U_r(\tau)} \bar{L} n^{-1} \left[\ln(F(\tau+r)-F(\tau)) - \ln\left(\frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d} \right) - \ln\left(n^{-\nu}d\right) \right] \right] \\
&= 0 \,.
\end{aligned}
$$

□

4.3 Results for $\bar{E}_n$

We deduce a similar result for the process $\bar{E}_n$. As the proofs show similarities to those of Section 4.2 their presentation will be shortened, while focusing on differences.

Lemma 4.7 Suppose that $\delta \in \mathbb{R}$ and $\bar{q} : (-\infty, \delta] \to (0, \infty)$ is a right-continuous and monotonically decreasing function. Let $u \in (\delta, \infty) \cup \{\infty\}$ and $\gamma < \delta$. If F is continuous on $(\gamma, \delta]$ then the inequality

$$
\mathbb{P}\left(\sup_{\gamma\le t\le\delta} \frac{\left|\bar{E}_n(t)-\bar{E}_n(u)\right|}{\bar{q}(t)} \ge \varepsilon \right) \le n^{-1}\varepsilon^{-2} \left[2\bar{q}(\delta)^{-2} \int\limits_{(\delta,u]} V(x)F(\mathrm{d}x) + \int\limits_{(\gamma,\delta]} \bar{q}(x-)^{-2} V(x)F(\mathrm{d}x) \right]
$$

holds for all $\varepsilon > 0$.

Proof. We follow the proof of Lemma 4.5 and use the same terminology for Ω_n, T_n, R_n, $a(\underline{x}_n)$, $b(\underline{x}_n)$, $c(\underline{x}_n)$, A, B and C. By noting that

$$
\bar{E}_n(t) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - m(X_i)) - E_n(t) = \frac{1}{n} \sum_{i=nF_n(t)+1}^{n} \left(Y_{[i:n]} - m(X_{i:n})\right)
$$

we get

$$\bar{\Delta}_n(t) := \bar{E}_n(t) - \bar{E}_n(u) = \frac{1}{n} \sum_{i=nF_n(t)+1}^{nF_n(u)} \left(Y_{[i:n]} - m(X_{i:n}) \right).$$

$\bar{\Delta}_n$ is a càdlàg process which is constant on $[X_{i:n}, X_{i+1:n})$ and with jumps in each $X_{i:n}$. Since $t \mapsto {}^1/_{\bar{q}(t)}$ is monotonically increasing, the process $\left\{ \left| \bar{\Delta}_n(t) \right| \bar{q}(t)^{-1}; \gamma \le t \le \delta \right\}$ is monotonically increasing on $[X_{i:n}, X_{i+1:n})$ and with jumps occurring at most in $X_{i:n}$ or at the discontinuity points of $\bar{q}$. In all discontinuity points of $\bar{q}$, which do not coincide with some $X_{i:n}$, the jump size is positive. All suprema of the process within $[X_{i:n}, X_{i+1:n})$ occur at the right boundary $X_{i+1:n}$. By the right continuity of the process, then for all $\omega \in \Omega_n$

$$\sup_{\gamma \le t \le \delta} \frac{\left| \bar{\Delta}_n(t) \right|}{\bar{q}(t)} = \max \left\{ \frac{\left| \bar{\Delta}_n(\delta) \right|}{\bar{q}(\delta)}, \max_{A \le k \le B} \frac{\left| \bar{\Delta}_n(X_{k:n}-) \right|}{\bar{q}(X_{k:n}-)} \right\}.$$

On the set $\Omega \setminus \Omega_n$ holds

$$\sup_{\gamma \le t \le \delta} \frac{\left| \bar{\Delta}_n(t) \right|}{\bar{q}(t)} = \frac{\left| \bar{\Delta}_n(\delta) \right|}{\bar{q}(\delta)}$$

and, hence,

$$\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{\left| \bar{\Delta}_n(t) \right|}{\bar{q}(t)} \ge \varepsilon \right) \le \mathbb{P}\left(\frac{\left| \bar{\Delta}_n(\delta) \right|}{\bar{q}(\delta)} \ge \varepsilon \right) + \mathbb{P}\left(\Omega_n \cap \left\{ \max_{A \le k \le B} \frac{\left| \bar{\Delta}_n(X_{k:n}-) \right|}{\bar{q}(X_{k:n}-)} \ge \varepsilon \right\} \right). \quad (4.13)$$

The first expression in (4.13) can be handled analogously to (4.10) and gives

$$\mathbb{P}\left(\frac{\left| \bar{\Delta}_n(\delta) \right|}{\bar{q}(\delta)} \ge \varepsilon \right) \le (\varepsilon \bar{q}(\delta))^{-2} n^{-1} \int_{(\delta, u]} V(x) F(\mathrm{d}x).$$

Now we handle the second summand on the right hand side of (4.13). As F is continuous on $(\gamma, \delta]$ we have $nF_n(X_{k:n}-) = k - 1$ if $A \le k \le B$ (cf. [Geo13, Remark 8.16]) and, therefore,

$$\bar{\Delta}_n(X_{k:n}-) = \frac{1}{n} \sum_{i=k}^{C} \left(Y_{[i:n]} - m(X_{i:n}) \right).$$

Define the mapping $\bar{H} : T_n \times \mathbb{R}^n \to \mathbb{R}$, with $\bar{H}(\underline{x}_n, \underline{y}_n) = \mathbb{1}_{\bar{M}}$ and

$$\bar{M} := \left\{ \max_{a(\underline{x}_n) \le k \le b(\underline{x}_n)} \frac{\left| \sum_{i=k}^{c(\underline{x}_n)} y_i - m(x_i) \right|}{\bar{q}(x_i -)} \ge n\varepsilon \right\} .$$

If $W_1, \ldots, W_n$ are independent random variables with $\mathbb{P} \circ (W_i)^{-1} = \mathbb{P}_{Y|X}(x_i, \cdot)$, then the same arguments used in the proof of Lemma 4.5 lead to

$$\begin{aligned}
&\mathbb{P}\left(\Omega_n \cap \left\{ \max_{A \le k \le B} \frac{\left|\bar{\Delta}_n(X_{k:n}-)\right|}{\bar{q}(X_{k:n}-)} \ge \varepsilon \right\} \right) \\
&= \int_{R_n} \mathbb{E}\left(\bar{H}\left(\underline{X}_n, \underline{Y}_n\right) | \underline{X}_n = \underline{x}_n \right) \left(\mathbb{P} \circ \underline{X}_n^{-1} \right) (\mathrm{d}\underline{x}_n) \\
&= \int_{R_n} \int \bar{H}(\underline{x}_n, \underline{y}_n) \bigotimes_{i=1}^{n} \mathbb{P}_{Y|X}(x_i, \mathrm{d}y_i) \left(\mathbb{P} \circ \underline{X}_n^{-1} \right) (\mathrm{d}\underline{x}_n) \\
&= \int_{R_n} \mathbb{P}\left(\max_{a \le k \le b} \frac{\left| \sum_{i=k}^{c} (W_i - m(x_i)) \right|}{\bar{q}(x_k -)} \ge n\varepsilon \right) \left(\mathbb{P} \circ \underline{X}_n^{-1} \right) (\mathrm{d}\underline{x}_n) .
\end{aligned}$$

Now set $W_i^* = W_{c-i+1}$ and $x_i^* := x_{c-i+1}$ and note that $\bar{q}(x_{c-b+1}^* -) \le \ldots \le \bar{q}(x_c^* -)$. By changing the order of summation and using the Hájek and Rényi inequality in A.2 it follows that

$$\begin{aligned}
&\mathbb{P}\left(\max_{a \le k \le b} \frac{\left| \sum_{i=k}^{c} (W_i - m(x_i)) \right|}{\bar{q}(x_k -)} \ge n\varepsilon \right) \\
&= \mathbb{P}\left(\max_{a \le k \le b} \frac{\left| \sum_{i=1}^{c-k+1} (W_i^* - m(x_i^*)) \right|}{\bar{q}(x_{c-k+1}^* -)} \ge n\varepsilon \right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}\left(\max_{c-b+1\le j\le c-a+1} \frac{\left|\sum_{i=1}^{j}(W_i^* - m(x_i^*))\right|}{\bar{q}(x_j^*-)} \ge n\varepsilon \right) \\
&\le (n\varepsilon)^{-2}\left\{ \bar{q}(x_{c-b+1}^*-)^{-2}\sum_{i=1}^{c-b+1} V(x_i^*) + \sum_{i=c-b+2}^{c-a+1} \bar{q}(x_i^*-)^{-2}V(x_i^*)\right\} \\
&= (n\varepsilon)^{-2}\left\{ \bar{q}(x_b-)^{-2}\sum_{i=b+1}^{c} V(x_i) + \sum_{i=a}^{b} \bar{q}(x_i-)^{-2}V(x_i)\right\} \\
&= (n\varepsilon)^{-2}\left\{ \bar{q}(x_b-)^{-2}\sum_{i=b}^{c} V(x_i) + \sum_{i=a}^{b-1} \bar{q}(x_i-)^{-2}V(x_i)\right\} .
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\mathbb{P}\left(\Omega_n \cap \left\{ \max_{A\le k\le B} \frac{\left|\bar{\Delta}_n(X_{k:n}-)\right|}{\bar{q}(X_{k:n}-)} \ge \varepsilon \right\} \right) \\
&\le \int_{R_n} (n\varepsilon)^{-2}\left\{ \bar{q}(x_b-)^{-2}\sum_{i=b+1}^{c} V(x_i) + \sum_{i=a}^{b} \bar{q}(x_i-)^{-2}V(x_i)\right\} \nu_n(\,\mathrm{d}\underline{x}_n) \\
&\le (n\varepsilon)^{-2}\left\{ \mathbb{E}\left(\bar{q}(\delta)^{-2}\sum_{i=1}^{n} \mathbb{1}_{X_i\in(\delta,u]}V(X_i)\right) + \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{1}_{X_i\in(\gamma,\delta]}\bar{q}(X_i-)^{-2}V(X_i)\right)\right\} \\
&= n^{-1}\varepsilon^{-2}\left\{ \bar{q}(\delta)^{-2}\int_{(\delta,u]} V(x)F(\,\mathrm{d}x) + \int_{(\gamma,\delta]} \bar{q}(x-)^{-2}V(x)F(\,\mathrm{d}x)\right\} . \qquad \square
\end{aligned}
$$

Corollary 4.8 If $\nu > 0$ and if $r > 0$ is chosen such that **(A2)-(A3)** holds for $U_r(\tau)$, then there exists a constant $C > 0$ such that

$$
\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le -t\le n^\nu r} \frac{\left|\bar{E}_n(\tau + n^{-\nu}t) - \bar{E}_n(\tau)\right|}{(-n^{-\nu}t)^{\frac{\nu+1}{2\nu}}} \ge \varepsilon \right) \le \varepsilon^{-2} d^{-\frac{1}{\nu}} C
$$

for all $\varepsilon > 0$ and $d > 0$. Moreover,

$$
\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le -t\le n^\nu r} \frac{\left|\bar{E}_n(\tau + n^{-\nu}t) - \bar{E}_n(\tau)\right|}{(-n^{-\nu}t)^{\lambda}} \ge \varepsilon \right) = 0
$$

for all $\varepsilon > 0$, $d > 0$ and $1/2 \le \lambda < \nu+1/2\nu$

Proof. We mostly follow the line of argument in the proof of Corollary 4.6. Fix $\varepsilon > 0$, $d > 0$ and for a given $\nu > 0$, let $\lambda \in \mathbb{R}$ be such that $1/2 \le \lambda \le \nu+1/2\nu$. Make the substitution $s = \tau + n^{-\nu} t$ and note that $\bar{q}(s) = (\tau - s)^{\lambda}$ is positive, continuous, and monotonically decreasing within $(-\infty, \tau - n^{-\nu} d]$. Thus, by Lemma 4.7 for sufficiently large $n \in \mathbb{N}$ it follows that

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{d \le -t \le n^{\nu} r} \frac{\left|\bar{E}_n(\tau + n^{-\nu} t) - \bar{E}_n(\tau)\right|}{(-n^{-\nu} t)^{\lambda}} \ge \varepsilon \right) \\
&= \mathbb{P}\left(\sup_{\tau - r \le s \le \tau - n^{-\nu} d} \frac{\left|\bar{E}_n(s) - \bar{E}_n(\tau)\right|}{(\tau - s)^{\lambda}} \ge \varepsilon \right) \\
&\le n^{-1} \varepsilon^{-2} \left[2(n^{-\nu} d)^{-2\lambda} \int\limits_{(\tau - n^{-\nu} d, \tau]} V(x) F(\mathrm{d}x) + \int\limits_{(\tau - r, \tau - n^{-\nu} d]} (\tau - x)^{-2\lambda} V(x) F(\mathrm{d}x) \right] \\
&\le n^{-1} \varepsilon^{-2} \left[2 n^{2\lambda\nu - \nu - 1} d^{-2\lambda + 1} \frac{\int_{(\tau - n^{-\nu} d, \tau]} V(x) F(\mathrm{d}x)}{F(\tau) - F(\tau - n^{-\nu} d)} \frac{F(\tau) - F(\tau - n^{-\nu} d)}{n^{-\nu} d} \right. \\
&\qquad \left. + \|V\|_{U_r(\tau)} \bar{L}^{\frac{\nu+1}{\nu}} \int\limits_{(\tau - r, \tau - n^{-\nu} d]} (F(\tau) - F(x))^{-2\lambda} F(\mathrm{d}x) \right].
\end{aligned}
$$

If $1/2 < \lambda \le \nu+1/2\nu$ choose $g(x) = (-2\lambda + 1)^{-1}(F(\tau) - x)^{-2\lambda+1}$, if otherwise $\lambda = 1/2$ set $g(x) = \ln(F(\tau) - x)$, and use the same arguments which where already used to show (4.11), to derive

$$
\begin{aligned}
&\int\limits_{(\tau - r, \tau - n^{-\nu} d]} (F(\tau) - F(x))^{-2\lambda} F(\mathrm{d}x) \\
&\quad = \begin{cases} \frac{1}{2\lambda - 1} \left[(F(\tau) - F(\tau - n^{-\nu} d))^{-2\lambda + 1} - (F(\tau) - F(\tau - r))^{-2\lambda + 1} \right] & \frac{1}{2} < \lambda \le \frac{\nu + 1}{2\nu} \\ \ln(F(\tau) - F(\tau - r)) - \ln(F(\tau) - F(\tau - n^{-\nu} d)) & \lambda = \frac{1}{2} \end{cases}
\end{aligned}
\tag{4.14}
$$

Finally,

$$
\limsup_{n \to \infty} \mathbb{P}\left(\sup_{d \le -t \le n^{\nu} r} \frac{\left|\bar{E}_n(\tau + n^{-\nu} t) - \bar{E}_n(\tau)\right|}{(-n^{-\nu} t)^{\frac{\nu+1}{2\nu}}} \ge \varepsilon \right)
$$

$$\leq \varepsilon^{-2} d^{-\frac{1}{\nu}} \left[2V(\tau-)F'_-(\tau) + \|V\|_{U_r(\tau)} \bar{L}^{\frac{\nu+1}{\nu}} \nu \left(F'_-(\tau)\right)^{-\frac{1}{\nu}} \right],$$

where the expression in the squared brackets is chosen to be the constant C. Otherwise, for $1/2 \leq \lambda < \nu+1/2\nu$ we have

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\leq -t\leq n^\nu r} \frac{\left|\bar{E}_n(\tau+n^{-\nu}t) - \bar{E}_n(\tau)\right|}{(-n^{-\nu}t)^\lambda} \geq \varepsilon \right) = 0\,. \qquad \square$$

4.4 Results for L_n

The process L_n can be written as a sum of a martingale and a compensator process. Due to this decomposition several inequalities will be proved.

Lemma 4.9 For each $n \in \mathbb{N}$ the process $(L_n(t))_{t\in\mathbb{R}\cup\{-\infty\}}$ can be decomposed into the sum

$$L_n(t) = I_n(t) + D_n(t)\,,$$

where $I_n(t)$ is a zero-mean and càdlàg square-integrable martingale indexed by $\mathbb{R}\cup\{-\infty\}$ and adapted to the sigma-field $\mathscr{F}_n(t)$, where $I_n(-\infty) = 0$, $\mathscr{F}_n(-\infty) := \{\emptyset, \Omega\}$ and

$$\mathscr{F}_n(t) := \sigma\left(\bigcup_{i=1}^n \sigma\left(\{\{X_i \leq r\}\,;\, r\leq t\}\right) \right), \quad t\in\mathbb{R}\,.$$

Furthermore,

$$D_n(t) = \int\limits_{(-\infty,t]} \frac{1-F_n(x-)}{1-F(x-)} m(x) F(\,\mathrm{d}x)\,, \quad t\in\mathbb{R}\,,$$

with $D_n(-\infty) = 0$, and

$$\mathbb{E}I_n^2(t) = \frac{1}{n} \int\limits_{(-\infty,t]} \frac{1-F(x)}{1-F(x-)} m^2(x) F(\,\mathrm{d}x)\,.$$

Proof. See [Fer09, Proposition 8.7]. $\square$

Lemma 4.10 Suppose that $\gamma \in \mathbb{R}$ and $q : [\gamma, \infty) \to (0, \infty)$ is a right-continuous and monotonically increasing function. If $u \in \{-\infty\} \cup (-\infty, \gamma]$, and $\delta > \gamma$ with $\delta \in T_F$, and if there is a dominating function $h(t)$ bounded on $[\gamma, \delta]$ with

$$\frac{|H(t) - H(u)|}{q(t)} \le h(t) \qquad \text{for all } t \in [\gamma, \delta]\,,$$

then the inequality

$$\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|(L_n(t) - H(t)) - (L_n(u) - H(u))|}{q(t)} \ge \varepsilon \right)$$
$$\le n^{-1} \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ q(\gamma)^{-2} \int_{(u,\gamma]} m^2(t) F(\mathrm{d}t) + \int_{(\gamma,\delta]} q(t)^{-2} m^2(t) F(\mathrm{d}t) + \frac{F(\delta)}{1 - F(\delta)} \left(\|h\|_\gamma^\delta \right)^2 \right\}$$

holds for all $\varepsilon > 0$.

Proof. Fix $\varepsilon > 0$. Use Lemma 4.9 to decompose $L_n(t) = I_n(t) + D_n(t)$ and estimate

$$\begin{aligned} &\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|(L_n(t) - H(t)) - (L_n(u) - H(u))|}{q(t)} \ge \varepsilon \right) \\ &\le \mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|I_n(t) - I_n(u)|}{q(t)} \ge \frac{\varepsilon}{2} \right) \\ &\qquad + \mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{|(D_n(t) - H(t)) - (D_n(u) - H(u))|}{q(t)} \ge \frac{\varepsilon}{2} \right) . \end{aligned} \tag{4.15}$$

For the first summand on the right hand side of (4.15), set $\hat{I}_n(t) := I_n(t) - I_n(u)$ and note that $\left(\hat{I}_n(t), \mathscr{F}_n(t)\right)_{t \ge u}$ is again a zero-mean and càdlàg square-integrable martingale with

$$\mathbb{E}\hat{I}_n^2(t) = \frac{1}{n} \int_{(u,t]} \frac{1 - F(x)}{1 - F(x-)} m^2(x) F(\mathrm{d}x)\,. \tag{4.16}$$

Moreover,

$$\mu(t) := \begin{cases} \mathbb{E}\hat{I}_n^2(\gamma)\,, & t < \gamma \\ \mathbb{E}\hat{I}_n^2(t)\,, & t \in [\gamma, \delta] \\ \mathbb{E}\hat{I}_n^2(\delta)\,, & t > \delta \end{cases}$$

is a measure-defining function for a measure with F-density $\mathbb{1}_{[\gamma,\delta]}{}^{(1-F)m^2}/_{n(1-F_-)}$. By Lemma 4.2, then

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\hat{I}_n(t)\right|}{q(t)}\ge\frac{\varepsilon}{2}\right)\\
&\le\left(\frac{\varepsilon}{2}\right)^{-2}\left\{q(\gamma)^{-2}\mu(\gamma)+\int\limits_{(\gamma,\delta]}q(t)^{-2}\mu(\,\mathrm{d}t)\right\}\\
&=\left(\frac{\varepsilon}{2}\right)^{-2}\left\{q(\gamma)^{-2}\frac{1}{n}\int\limits_{(u,\gamma]}\frac{1-F(t)}{1-F(t-)}m^2(t)F(\,\mathrm{d}t)+\frac{1}{n}\int\limits_{(\gamma,\delta]}q(t)^{-2}\frac{1-F(t)}{1-F(t-)}m^2(t)F(\,\mathrm{d}t)\right\}\\
&\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\left\{q(\gamma)^{-2}\int\limits_{(u,\gamma]}m^2(t)F(\,\mathrm{d}t)+\int\limits_{(\gamma,\delta]}q(t)^{-2}m^2(t)F(\,\mathrm{d}t)\right\}\,.
\end{aligned}$$

Inspecting the second term in (4.15) and noting that

$$\begin{aligned}
&\sup_{\gamma\le t\le\delta}\frac{|(D_n(t)-H(t))-(D_n(u)-H(u))|}{q(t)}\\
&=\sup_{\gamma\le t\le\delta}\frac{\left|\int\limits_{(u,t]}\frac{F(x-)-F_n(x-)}{1-F(x-)}m(x)F(\,\mathrm{d}x)\right|}{q(t)}\\
&\le\sup_{\gamma\le t\le\delta}\left\|\frac{F_n-F}{1-F}\right\|_u^t\frac{|H(t)-H(u)|}{q(t)}\\
&\le\left\|\frac{F_n-F}{1-F}\right\|_u^\delta\|h\|_\gamma^\delta\,,
\end{aligned}$$

yields

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|(D_n(t)-H(t))-(D_n(u)-H(u))|}{q(t)}\ge\frac{\varepsilon}{2}\right)\\
&\le\mathbb{P}\left(\|h\|_\gamma^\delta\sup_{u\le t\le\delta}\left|\frac{F_n(t)-F(t)}{1-F(t)}\right|\ge\frac{\varepsilon}{2}\right)\\
&\overset{(a)}{\le}\left(\frac{\varepsilon}{2}\right)^{-2}\left(\|h\|_\gamma^\delta\right)^2\mathbb{E}\left(\frac{F_n(\delta)-F(\delta)}{1-F(\delta)}\right)^2\\
&\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\frac{F(\delta)}{1-F(\delta)}\left(\|h\|_\gamma^\delta\right)^2\,,
\end{aligned}$$

where we used in (a) the submartingale property of $({}^{F_n-F}/_{1-F})^2$ (see [Kou02, p. 68]) together with Doob's inequality (see [SW86, p. 874]). $\square$

Corollary 4.11 If $\nu > 0$ and if $r > 0$ such that **(A2)** holds for $U_r(\tau)$ with $U_r(\tau) \subseteq T_F$ and $\|m\|_{U_r(\tau)\setminus\{\tau\}} < \infty$, then there exists a constant $C > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le t\le n^\nu r} \frac{|(L_n(\tau+n^{-\nu}t)-H(\tau+n^{-\nu}t))-(L_n(\tau)-H(\tau))|}{(n^{-\nu}t)^{\frac{\nu+1}{2\nu}}} \ge \varepsilon \right) \le C\left(\frac{\varepsilon}{2}\right)^{-2} d^{-\frac{1}{\nu}}$$

for all $\varepsilon > 0$ and $d > 0$. Moreover,

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le t\le n^\nu r} \frac{|(L_n(\tau+n^{-\nu}t)-H(\tau+n^{-\nu}t))-(L_n(\tau)-H(\tau))|}{(n^{-\nu}t)^{\lambda}} \ge \varepsilon \right) = 0$$

for all $\varepsilon > 0$, $d > 0$ and ${}^1/_2 \le \lambda < {}^{\nu+1}/_{2\nu}$.

Proof. Fix $\varepsilon > 0$ and $d > 0$. For a given $\nu > 0$, let $\lambda \in \mathbb{R}$ be such that ${}^1/_2 \le \lambda \le {}^{\nu+1}/_{2\nu}$. Substitute $\tau + n^{-\nu}t$ by s and note that

$$\begin{aligned} &\mathbb{P}\left(\sup_{d\le t\le n^\nu r} \frac{|(L_n(\tau+n^{-\nu}t)-H(\tau+n^{-\nu}t))-(L_n(\tau)-H(\tau))|}{(n^{-\nu}t)^{\lambda}} \ge \varepsilon \right) \\ &\qquad = \mathbb{P}\left(\sup_{\tau+n^{-\nu}d\le s\le \tau+r} \frac{|(L_n(s)-H(s))-(L_n(\tau)-H(\tau))|}{(s-\tau)^{\lambda}} \ge \varepsilon \right) . \end{aligned}$$

The properties for m and assumption **(A2)** provides that for all $s \in [\tau + n^{-\nu}d, \tau + r]$,

$$\frac{|H(s)-H(\tau)|}{(s-\tau)^\lambda} = \frac{\left| \int_{(\tau,s]} m(x)F(\mathrm{d}x) \right|}{(s-\tau)^\lambda} \le \|m\|_{U_r(\tau)\setminus\{\tau\}}\, \bar{L}(s-\tau)^{1-\lambda} .$$

Define $h(s) := (s-\tau)^{1-\lambda}$ and assume that ${}^1/_2 < \lambda \le {}^{\nu+1}/_{2\nu}$. Applying Lemma 4.10 in (a) below and in (b) below the change-of-variable theorem in the same way as in (4.11) yields

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\tau+n^{-\nu}d\le s\le\tau+r} \frac{|(L_n(s)-H(s))-(L_n(\tau)-H(\tau))|}{(s-\tau)^\lambda} \ge \varepsilon \right) \\
&\overset{(a)}{\le} \limsup_{n\to\infty} n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2} \left\{ \left(n^{-\nu}d\right)^{-2\lambda} \int\limits_{(\tau,\tau+n^{-\nu}d]} m^2(x)\,F(\,\mathrm{d}x) \right. \\
&\quad \left. + \int\limits_{(\tau+n^{-\nu}d,\tau+r]} (x-\tau)^{-2\lambda} m^2(x) F(\,\mathrm{d}x) + \frac{F(\tau+r)}{1-F(\tau+r)} \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L}^2 \left(\|h\|^{\tau+r}_{\tau+n^{-\nu}d}\right)^2 \right\} \\
&\le \limsup_{n\to\infty} \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ n^{2\lambda\nu-\nu-1} d^{-2\lambda+1} \frac{\int_{(\tau,\tau+n^{-\nu}d]} m^2(x)\,F(\,\mathrm{d}x)}{F(\tau+n^{-\nu}d)-F(\tau)} \frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d} \right. \\
&\quad + n^{-1} \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L}^{2\lambda} \int\limits_{(\tau+n^{-\nu}d,\tau+r]} (F(x)-F(\tau))^{-2\lambda} F(\,\mathrm{d}x) \\
&\quad \left. + n^{-1} \frac{F(\tau+r)}{1-F(\tau+r)} \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L}^2 \left(\|h\|^{\tau+r}_{\tau+n^{-\nu}d}\right)^2 \right\} \\
&\overset{(b)}{=} \limsup_{n\to\infty} \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ n^{2\lambda\nu-\nu-1} d^{-2\lambda+1} \frac{\int_{(\tau,\tau+n^{-\nu}d]} m^2(x)\,F(\,\mathrm{d}x)}{F(\tau+n^{-\nu}d)-F(\tau)} \frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d} \right. \\
&\quad + \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L}^{2\lambda} \frac{1}{2\lambda-1} \left[n^{2\lambda\nu-\nu-1} d^{-2\lambda+1} \left(\frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d} \right)^{-2\lambda+1} \right. \\
&\quad \left. - n^{-1} (F(\tau+r)-F(\tau))^{-2\lambda+1} \right] \\
&\quad \left. + n^{-1} \frac{F(\tau+r)}{1-F(\tau+r)} \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L}^2 \left(\|h\|^{\tau+r}_{\tau+n^{-\nu}d}\right)^2 \right\} .
\end{aligned}$$

Furthermore, since $\lambda < {}^{2\nu+1}/_{2\nu}$,

$$n^{-1}\left(\|h\|^{\tau+r}_{\tau+n^{-\nu}d}\right)^2 = \begin{cases} n^{-1} r^{2-2\lambda}, & \text{if } \lambda < 1 \\ n^{-1}, & \text{if } \lambda = 1 \\ n^{2\lambda\nu-2\nu-1} d^{2-2\lambda}, & \text{if } \lambda > 1 \end{cases} \longrightarrow 0\,,$$

as $n \to \infty$. If $\lambda = {}^{\nu+1}/_{2\nu}$, then

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le t\le n^{\nu}r} \frac{|(L_n(\tau+n^{-\nu}t)-H(\tau+n^{-\nu}t))-(L_n(\tau)-H(\tau))|}{(n^{-\nu}t)^{\frac{\nu+1}{2\nu}}} \ge \varepsilon \right)$$
$$\le \left(\frac{\varepsilon}{2}\right)^{-2} d^{-\frac{1}{\nu}} \left\{ m^2(\tau+)F'_+(\tau) + \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L}^{\frac{\nu+1}{\nu}} \nu \left(F'_+(\tau)\right)^{-\frac{1}{\nu}} \right\} ,$$

where the expression in braces is a suitable positive constant C. Otherwise, if we have ${}^{1}/_{2} < \lambda < {}^{\nu+1}/_{2\nu}$, then

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le t\le n^{\nu}r} \frac{|(L_n(\tau+n^{-\nu}t)-H(\tau+n^{-\nu}t))-(L_n(\tau)-H(\tau))|}{(n^{-\nu}t)^{\lambda}} \ge \varepsilon \right) = 0 .$$

As in the proof of Corollary 4.6, treat the case $\lambda = {}^{1}/_{2}$ separately and find

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{d\le t\le n^{\nu}r} \frac{|(L_n(\tau+n^{-\nu}t)-H(\tau+n^{-\nu}t))-(L_n(\tau)-H(\tau))|}{(n^{-\nu}t)^{\frac{1}{2}}} \ge \varepsilon \right)$$
$$\le \limsup_{n\to\infty} \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ n^{-1} \frac{\int_{(\tau,\tau+n^{-\nu}d]} m^2(x)\,F(\mathrm{d}x)}{F(\tau+n^{-\nu}d)-F(\tau)} \frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d} \right.$$
$$+ \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L} n^{-1} \left[\ln(F(\tau+r)-F(\tau)) - \ln\left(\frac{F(\tau+n^{-\nu}d)-F(\tau)}{n^{-\nu}d}\right) - \ln\left(n^{-\nu}d\right) \right]$$
$$\left. + n^{-1} \frac{F(\tau+r)}{1-F(\tau+r)} \|m\|^2_{U_r(\tau)\setminus\{\tau\}} \bar{L}^2 r \right\}$$
$$= 0 . \qquad \square$$

Lemma 4.12 If $\|m\| < \infty$ then for each interval $[\gamma,\delta] \subseteq T_F$ there exists some constant $C > 0$ such that

$$\mathbb{P}\left(\sup_{\gamma\le t\le\delta} \frac{|L_n(t)-H(t)|}{F(t)} \ge \varepsilon \right) \le n^{-1} \left(\frac{\varepsilon}{2}\right)^{-2} \|m\|^2 \left\{ 2F(\gamma)^{-1} + C \right\}$$

for all $\varepsilon > 0$.

Proof. First imitate the proof of Lemma 4.10. Then the inequality

$$\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|L_n(t)-H(t)|}{F(t)}\ge\varepsilon\right) \le \mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|I_n(t)|}{F(t)}\ge\frac{\varepsilon}{2}\right)+\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|D_n(t)-H(t)|}{F(t)}\ge\frac{\varepsilon}{2}\right) \tag{4.17}$$

holds, and using Lemma 4.2 for the first term yields

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|I_n(t)|}{F(t)}\ge\frac{\varepsilon}{2}\right)\\
&\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\left\{F(\gamma)^{-2}\int\limits_{(-\infty,\gamma]} m^2(x)\,F(\,\mathrm{d}x)+\int\limits_{(\gamma,\delta]}F(x)^{-2}m^2(x)F(\,\mathrm{d}x)\right\}\\
&\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|m\|^2\left\{F(\gamma)^{-1}+\int\limits_{(\gamma,\delta]}F(x)^{-2}F(\,\mathrm{d}x)\right\}\\
&\overset{(a)}{\le} n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|m\|^2\,2F(\gamma)^{-1}\,.
\end{aligned}$$

To verify (a), use the quantile transformation $X = F^{-1}(U)$ for some uniform random variable U. By the inequalities a) and e) in [Wit85, Lemma 1.17] it follows that

$$\int\limits_{(\gamma,\delta]}F(x)^{-2}F(\,\mathrm{d}x)=\int\limits_{\{F(\gamma)<U\le F(\delta)\}}F(F^{-1}(U))^{-2}\,\mathrm{d}\mathbb{P}\le\int\limits_{F(\gamma)}^{F(\delta)}u^{-2}\,\mathrm{d}u\le F(\gamma)^{-1}\,.$$

We now have to modify the estimate for the second summand on the right hand side of (4.17). For every $\eta\in(1/2,1)$ we can assert that

$$\begin{aligned}
&\frac{|D_n(t)-H(t)|}{F(t)}\\
&=\left|\int\limits_{(-\infty,t]}\frac{F(x-)-F_n(x-)}{1-F(x-)}m(x)F(\,\mathrm{d}x)\right|\frac{1}{F(t)}
\end{aligned}$$

$$= \left| \int\limits_{(-\infty,t]} \frac{(F(x-)-F_n(x-))\,/\,(1-F(x-))}{(1-F(x-))^{-\eta}} \frac{m(x)}{(1-F(x-))^{\eta}} F(\,\mathrm{d}x) \right| \frac{1}{F(t)}$$

$$\leq \|m\| \sup_{-\infty<x\leq t} \frac{|(F_n(x-)-F(x-))\,/\,(1-F(x-))|}{(1-F(x-))^{-\eta}} \frac{\int_{(-\infty,t]} \frac{1}{(1-F(x-))^{\eta}} F(\,\mathrm{d}x)}{F(t)} \,. \tag{4.18}$$

After making the quantile transformation $X = F^{-1}(U)$, we observe that $F(F^{-1}(u)-) \leq u$ for all $u \in [0,1]$ (see [Wit85, Lemma 1.17 e)]). Since $1-\eta \in (0, 1/2)$ and $t \in T_F$ for all $t \in [\gamma, \delta]$, we have $(1-F(t))^{1-\eta} \geq 1-F(t)$ and, hence,

$$\int\limits_{(-\infty,t]} \frac{1}{(1-F(x-))^{\eta}} F(\,\mathrm{d}x) \leq \int\limits_{0}^{F(t)} \frac{1}{(1-u)^{\eta}} \,\mathrm{d}u = \frac{1}{1-\eta}\left(1-(1-F(t))^{1-\eta}\right) \leq \frac{F(t)}{1-\eta} \,.$$

Note that $\frac{F_n(x)-F(x)}{1-F(x)}$ is a zero-mean càdlàg martingale on $(-\infty,\delta]$ (cf. [Kou02, p. 68]) with

$$\mathbb{E}\left(\frac{F_n(x)-F(x)}{1-F(x)}\right)^2 = \frac{\mathrm{Var}(F_n(x))}{(1-F(x))^2} = n^{-1}\frac{F(x)}{1-F(x)} \,.$$

Furthermore, $(1-F(x))^{-\eta}$ is positive, non-decreasing and right-continuous on $(-\infty,\delta]$. We successively apply (4.18), continuity of $\mathbb{P}$ from below, A.7 and Lemma 4.2, and get

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma\leq t\leq\delta} \frac{|D_n(t)-H(t)|}{F(t)} \geq \frac{\varepsilon}{2}\right) \\
&\leq \mathbb{P}\left(\sup_{\gamma\leq t\leq\delta} \sup_{-\infty<x\leq t} \frac{|(F_n(x-)-F(x-))\,/\,(1-F(x-))|}{(1-F(x-))^{-\eta}} \geq \frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right) \\
&\leq \mathbb{P}\left(\sup_{-\infty<x\leq\delta} \frac{|(F_n(x-)-F(x-))\,/\,(1-F(x-))|}{(1-F(x-))^{-\eta}} \geq \frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right) \\
&= \lim_{w\downarrow-\infty} \mathbb{P}\left(\sup_{w<x\leq\delta} \frac{|(F_n(x-)-F(x-))\,/\,(1-F(x-))|}{(1-F(x-))^{-\eta}} \geq \frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right) \\
&= \lim_{w\downarrow-\infty} \mathbb{P}\left(\sup_{w\leq x<\delta} \frac{|(F_n(x)-F(x))\,/\,(1-F(x))|}{(1-F(x))^{-\eta}} \geq \frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right) \\
&= \lim_{w\downarrow-\infty} \lim_{\delta_k\uparrow\delta} \mathbb{P}\left(\sup_{w\leq x\leq\delta_k} \frac{|(F_n(x)-F(x))\,/\,(1-F(x))|}{(1-F(x))^{-\eta}} \geq \frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)
\end{aligned}$$

$$\leq \left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2} \lim_{w\downarrow-\infty}\lim_{\delta_k\uparrow\delta}\Bigg\{(1-F(w))^{2\eta}n^{-1}\frac{F(w)}{1-F(w)} + \int\limits_{(w,\delta_k]}(1-F(x))^{2\eta}\,\mathrm{d}\left(n^{-1}\frac{F}{1-F}\right)\Bigg\}$$

$$\stackrel{(b)}{=} \left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2} n^{-1}\lim_{w\downarrow-\infty}\lim_{\delta_k\uparrow\delta}\Bigg\{(1-F(w))^{2\eta-1}F(w) + \int\limits_{(w,\delta_k]}\frac{(1-F(x))^{2\eta-1}}{1-F(x-)}F(\mathrm{d}x)\Bigg\}$$

$$\stackrel{(c)}{\leq} \left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2} n^{-1}\lim_{w\downarrow-\infty}\lim_{\delta_k\uparrow\delta}\left\{(1-F(w))^{2\eta-1}F(w) + \frac{1}{2\eta-1}(F(\delta_k)-F(w))^{2\eta-1}\right\}$$

$$= \left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2} n^{-1}\left\{\frac{1}{2\eta-1}F(\delta-)^{2\eta-1}\right\}$$

$$\leq n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\frac{\|m\|^2}{(1-\eta)^2(2\eta-1)}\,.$$

In (b) we applied the Integration by parts theorem for Lebesgue-Stieltjes integrals, cf. [HS75, Theorem 21.67] and [SW86, page 868-869], to see that the measure corresponding to $F/n(1-F)$ is a measure with F-density $1/n(1-F)(1-F_-)$. To check (c) use again a quantile transformation $X = F^{-1}(U)$ for some uniform random variable U. Observe that $F(F^{-1}(U)-) \leq U$, see [Wit85, Lemma 1.17], and that η was chosen so that $(2\eta-2)\in(-1,0)$. Hence, $\left(1-F(F^{-1}(U)-)\right)^{2\eta-2} \leq (1-U)^{2\eta-2}$. Moreover, the function $g(x) = x^{2\eta-1}$ is concave and for $0 \leq x < y \leq 1$ we have $g(y-x) = g(\frac{x}{y}0 + \frac{y-x}{y}y) \geq g(y) - \frac{x}{y}g(y)$ and $g(x) = g(\frac{y-x}{y}0 + \frac{x}{y}y) \geq \frac{x}{y}g(y)$. Consequently, $g(y-x) \geq g(y) - g(x)$, and we observe that

$$\int\limits_{(w,\delta_k]}\frac{(1-F(x))^{2\eta-1}}{1-F(x-)}F(\mathrm{d}x)$$

$$\leq \int\limits_{(w,\delta_k]}(1-F(x-))^{2\eta-2}F(\mathrm{d}x)$$

$$= \int\limits_{\{F(w)<U\leq F(\delta_k)\}}(1-F(F^{-1}(U)-))^{2\eta-2}\,\mathrm{d}\mathbb{P}$$

$$\leq \int\limits_{F(w)}^{F(\delta)}(1-u)^{2\eta-2}\,\mathrm{d}u$$

$$
\begin{aligned}
&= \frac{1}{2\eta-1}\left[(1-F(w))^{2\eta-1}-(1-F(\delta_k))^{2\eta-1}\right] \\
&= \frac{1}{2\eta-1}\left[g(1-F(w))-g(1-F(\delta_k))\right] \\
&\le \frac{1}{2\eta-1}(F(\delta_k)-F(w))^{2\eta-1} .
\end{aligned}
$$

Finally,

$$
\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{|L_n(t)-H(t)|}{F(t)}\ge\varepsilon\right)\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|m\|^2\left\{2F(\gamma)^{-1}+C\right\}
$$

holds where $C=\left((1-\eta)^2(2\eta-1)\right)^{-1}>0$. □

4.5 Results for $\bar{L}_n$

In contrast to L_n, the process $\bar{L}_n$ can be written as a sum of a backwards martingale and a compensator. Therefore, we will reformulate the results of Section 4.4. Again, as the proofs show similarities their presentation will be shortened, while focusing on differences.

Lemma 4.13 For each $n\in\mathbb{N}$ the process $(\bar{L}_n(t))_{t\in\mathbb{R}\cup\{+\infty\}}$ can be decomposed into the sum

$$
\bar{L}_n(t)=\bar{I}_n(t)+\bar{D}_n(t) ,
$$

where $\bar{I}_n(t)$ is a zero-mean and càdlàg square-integrable backwards martingale indexed by $\mathbb{R}\cup\{+\infty\}$ and adapted to the sigma-field $\bar{\mathscr{F}}_n(t)$, where $\bar{I}_n(+\infty)=0$, $\bar{\mathscr{F}}_n(+\infty)=\{\emptyset,\Omega\}$ and

$$
\bar{\mathscr{F}}_n(t):=\sigma\left(\bigcup_{i=1}^{n}\sigma\left(\{\{X_i>r\}\,;\,r\ge t\}\right)\right),\quad t\in\mathbb{R} .
$$

Furthermore,

$$
\bar{D}_n(t)=\int\limits_{(t,\infty)}\frac{F_n(x)}{F(x)}m(x)F(\mathrm{d}x) ,\quad t\in\mathbb{R} ,
$$

with $\bar{D}_n(+\infty) = 0$, and

$$\mathbb{E}\bar{I}_n^2(t) = \frac{1}{n} \int\limits_{(t,\infty)} \frac{F(x-)}{F(x)} m^2(x) F(\mathrm{d}x)\,.$$

Proof. The result is formulated in [Til07, Lemma 4.26] without proof. However, the proof may be adjusted to the backwards case without much effort. □

Lemma 4.14 Suppose that $\delta \in \mathbb{R}$ and $\bar{q} : (-\infty, \delta] \to (0, \infty)$ is a right-continuous and monotonically decreasing function. If $u \in [\delta, \infty) \cup \{\infty\}$, and $\gamma < \delta$ with $\gamma \in T_F$, and if there is a dominating function $\bar{h}(t)$ bounded on $[\gamma, \delta]$ with

$$\frac{\left|\bar{H}(t) - \bar{H}(u)\right|}{\bar{q}(t)} \leq \bar{h}(t) \qquad \text{for all } t \in [\gamma, \delta]\,,$$

then the inequality

$$\begin{aligned} &\mathbb{P}\left(\sup_{\gamma \leq t \leq \delta} \frac{\left|(\bar{L}_n(t) - \bar{H}(t)) - (\bar{L}_n(u) - \bar{H}(u))\right|}{\bar{q}(t)} \geq \varepsilon \right) \\ &\leq n^{-1} \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ \bar{q}(\delta)^{-2} \int\limits_{(\delta,u]} m^2(t) F(\mathrm{d}t) + \int\limits_{(\gamma,\delta]} \bar{q}(t-)^{-2} m^2(t) F(\mathrm{d}t) + \frac{1 - F(\gamma)}{F(\gamma)} \left\|\bar{h}^2\right\|_\gamma^\delta \right\} \end{aligned}$$

holds for all $\varepsilon > 0$.

Proof. Similar to the proof of Lemma 4.10, use Lemma 4.13 to show that

$$\begin{aligned} &\mathbb{P}\left(\sup_{\gamma \leq t \leq \delta} \frac{\left|\left(\bar{L}_n(t) - \bar{H}(t)\right) - \left(\bar{L}_n(u) - \bar{H}(u)\right)\right|}{\bar{q}(t)} > \varepsilon \right) \\ &\leq \mathbb{P}\left(\sup_{\gamma \leq t \leq \delta} \frac{\left|\bar{I}_n(t) - \bar{I}_n(u)\right|}{\bar{q}(t)} > \frac{\varepsilon}{2} \right) \\ &\qquad + \mathbb{P}\left(\sup_{\gamma \leq t \leq \delta} \frac{\left|\left(\bar{D}_n(t) - \bar{H}(t)\right) - \left(\bar{D}_n(u) - \bar{H}(u)\right)\right|}{\bar{q}(t)} > \frac{\varepsilon}{2} \right)\,. \end{aligned} \tag{4.19}$$

Set $\check{I}_n(t) := \bar{I}_n(t) - \bar{I}_n(u)$ and note that $\left(\check{I}_n(t), \bar{\mathscr{F}}_n(t)\right)_{t \leq u}$ is again a zero-mean càdlàg and square-integrable backwards martingale with

$$\mathbb{E}\check{I}_n^2(t) = \frac{1}{n} \int\limits_{(t,u]} \frac{F(x-)}{F(x)} m^2(x) F(\mathrm{d}x)\,. \tag{4.20}$$

For intervals $(a,b] \subseteq [\gamma,\delta]$ we use the fact that

$$\begin{aligned}
\mu((a,b]) :=& -\mathbb{E}\check{I}_n^2(b) + \mathbb{E}\check{I}_n^2(a) \\
=& \frac{1}{n}\left(\int\limits_{(a,u]} \frac{F(t-)}{F(t)} m^2(t) F(\mathrm{d}t) - \int\limits_{(b,u]} \frac{F(t-)}{F(t)} m^2(t) F(\mathrm{d}t) \right) \\
=& \frac{1}{n}\left(\int\limits_{(a,b]} \frac{F(t-)}{F(t)} m^2(t) F(\mathrm{d}t) \right),
\end{aligned}$$

which means that μ is a measure with F-density $\mathbb{1}_{[\gamma,\delta]} {}^{F_- m^2}/_{nF}$. We proceed analogously to the proof of Lemma 4.10 and apply the Birnbaum-Marshall type inequality for backwards martingales in Lemma 4.4:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma \le t \le \delta} \frac{\left|\check{I}_n(t)\right|}{\bar{q}(t)} > \frac{\varepsilon}{2} \right) \\
&\le \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ \bar{q}(\delta)^{-2}\, \mathbb{E}\check{I}_n^2(\delta) + \int\limits_{(\gamma,\delta]} \bar{q}(t-)^{-2} \left(-\mathbb{E}\check{I}_n^2(\mathrm{d}t)\right) \right\} \\
&= \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ \bar{q}(\delta)^{-2} \frac{1}{n} \int\limits_{(\delta,u]} \frac{F(t-)}{F(t)} m^2(t) F(\mathrm{d}t) + \frac{1}{n} \int\limits_{(\gamma,\delta]} \bar{q}(t-)^{-2} \frac{F(t-)}{F(t)} m^2(t) F(\mathrm{d}t) \right\} \\
&\overset{(a)}{\le} n^{-1} \left(\frac{\varepsilon}{2}\right)^{-2} \left\{ \bar{q}(\delta)^{-2} \int\limits_{(\delta,u]} m^2(t) F(\mathrm{d}t) + \int\limits_{(\gamma,\delta]} \bar{q}(t-)^{-2} m^2(t) F(\mathrm{d}t) \right\}.
\end{aligned}$$

In Inequality (a) we used ${}^{F_-}/_{F} \le 1$. To estimate the second summand on the right hand side of Equation (4.19), note that

$$\frac{\left|\left(\bar{D}_n(t) - \bar{H}(t)\right) - \left(\bar{D}_n(u) - \bar{H}(u)\right)\right|}{\bar{q}(t)} = \frac{\left| \int\limits_{(t,u]} \frac{F_n(x) - F(x)}{F(x)} m(x) F(\mathrm{d}x) \right|}{\bar{q}(t)} \le \left\| \frac{F_n - F}{F} \right\|_t^u \bar{h}(t)$$

and, hence,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\left(\bar{D}_n(t)-\bar{H}(t)\right)-\left(\bar{D}_n(u)-\bar{H}(u)\right)\right|}{\bar{q}(t)}>\frac{\varepsilon}{2}\right)\\
&\le\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\left\|\frac{F_n-F}{F}\right\|_t^u\bar{h}(t)>\frac{\varepsilon}{2}\right)\\
&\le\mathbb{P}\left(\left\|\bar{h}\right\|_\gamma^\delta\sup_{\gamma\le t\le u}\left|\frac{F_n(t)-F(t)}{F(t)}\right|>\frac{\varepsilon}{2}\right)\\
&\overset{(b)}{=}\mathbb{P}\left(\left\|\bar{h}\right\|_\gamma^\delta\sup_{-u\le t\le-\gamma}\left|\frac{F_n(-t)-F(-t)}{F(-t)}\right|>\frac{\varepsilon}{2}\right)\\
&\overset{(c)}{\le}\left(\frac{\varepsilon}{2}\right)^{-2}\left\|\bar{h}^2\right\|_\gamma^\delta\,\mathbb{E}\left(\frac{F_n(\gamma)-F(\gamma)}{F(\gamma)}\right)^2\\
&\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\frac{1-F(\gamma)}{F(\gamma)}\left\|\bar{h}^2\right\|_\gamma^\delta\ .
\end{aligned}$$

In (b) we used that F_n-F/F is a backwards martingale (cf. [Kou02, section 2.4.3]) and, therefore, $F_n(-t)-F(-t)/F(-t)$ is a martingale. We finally used Doob's inequality in (c). □

Corollary 4.15 If $\nu>0$ and if $r>0$ such that **(A2)** holds for $U_r(\tau)$ with $U_r(\tau)\subseteq T_F$ and $\|m\|_{U_r(\tau)\setminus\{\tau\}}<\infty$, then there exists a constant $C>0$ such that

$$\limsup_{n\to\infty}\mathbb{P}\left(\sup_{d\le-t\le n^\nu r}\frac{\left|\left(\bar{L}_n(\tau+n^{-\nu}t)-\bar{H}(\tau+n^{-\nu}t)\right)-\left(\bar{L}_n(\tau)-\bar{H}(\tau)\right)\right|}{(-n^{-\nu}t)^{\frac{\nu+1}{2\nu}}}\ge\varepsilon\right)\le C\left(\frac{\varepsilon}{2}\right)^{-2}d^{-\frac{1}{\nu}}$$

for all $\varepsilon>0$ and $d>0$. Moreover,

$$\limsup_{n\to\infty}\mathbb{P}\left(\sup_{d\le-t\le n^\nu r}\frac{\left|\left(\bar{L}_n(\tau+n^{-\nu}t)-\bar{H}(\tau+n^{-\nu}t)\right)-\left(\bar{L}_n(\tau)-\bar{H}(\tau)\right)\right|}{(-n^{-\nu}t)^{\lambda}}\ge\varepsilon\right)=0$$

for all $\varepsilon>0$, $d>0$ and $1/2\le\lambda<\nu+1/2\nu$.

Proof. Fix $\varepsilon>0$, $d>0$ and $\nu>0$. Let $\lambda\in\mathbb{R}$ be such that $1/2\le\lambda\le\nu+1/2\nu$. Again substitute $\tau+n^{-\nu}t$ by s. Due to the assumptions for F and m, for all $s\in[\tau-r,\tau-n^{-\nu}d]$ holds

$$\frac{\left|\bar{H}(s)-\bar{H}(\tau)\right|}{(\tau-s)^\lambda}\le\|m\|_{U_r(\tau)\setminus\{\tau\}}\,\bar{L}(\tau-s)^{1-\lambda}=:\|m\|_{U_r(\tau)\setminus\{\tau\}}\,\bar{L}\bar{h}(s)\ .$$

Analysis similar to that in the proof of Corollary 4.11, application of Lemma 4.14, and the change of variables theorem (cf. with (4.14)) yields

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{P}\left(\sup_{d\le -t\le n^{\nu}r}\frac{\left|\left(\bar{L}_n(\tau+n^{-\nu}t)-\bar{H}(\tau+n^{-\nu}t)\right)-\left(\bar{L}_n(\tau)-\bar{H}(\tau)\right)\right|}{(-n^{-\nu}t)^{\lambda}}\ge\varepsilon\right)\\
&=\limsup_{n\to\infty}\mathbb{P}\left(\sup_{\tau-r\le s\le\tau-n^{-\nu}d}\frac{\left|(\bar{L}_n(s)-\bar{H}(s))-(\bar{L}_n(\tau)-\bar{H}(\tau))\right|}{(\tau-s)^{\lambda}}\ge\varepsilon\right)\\
&\le\limsup_{n\to\infty}n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\Bigg\{\left(n^{-\nu}d\right)^{-2\lambda}\int\limits_{(\tau-n^{-\nu}d,\tau]}m^2(x)F(\mathrm{d}x)\\
&\quad+\int\limits_{(\tau-r,\tau-n^{-\nu}d]}(\tau-x)^{-2\lambda}m^2(x)F(\mathrm{d}x)+\frac{1-F(\tau-r)}{F(\tau-r)}\|m\|^2_{U_r(\tau)\setminus\{\tau\}}\bar{L}^2\left(\left\|\bar{h}\right\|_{\tau-r}^{\tau-n^{-\nu}d}\right)^2\Bigg\}\\
&\le\limsup_{n\to\infty}\left(\frac{\varepsilon}{2}\right)^{-2}\Bigg\{n^{2\lambda\nu-\nu-1}d^{-2\lambda+1}\frac{\int_{(\tau-n^{-\nu}d,\tau]}m^2(x)F(\mathrm{d}x)}{F(\tau)-F(\tau-n^{-\nu}d)}\frac{F(\tau)-F(\tau-n^{-\nu}d)}{n^{-\nu}d}\\
&\quad+\|m\|^2_{U_r(\tau)\setminus\{\tau\}}\bar{L}^{2\lambda}\frac{1}{2\lambda-1}\left[n^{2\lambda\nu-\nu-1}d^{-2\lambda+1}\left(\frac{F(\tau)-F(\tau-n^{-\nu}d)}{n^{-\nu}d}\right)^{-2\lambda+1}\right.\\
&\quad\left.-n^{-1}(F(\tau)-F(\tau-r))^{-2\lambda+1}\right]\\
&\quad+n^{-1}\frac{1-F(\tau-r)}{F(\tau-r)}\|m\|^2_{U_r(\tau)\setminus\{\tau\}}\bar{L}^2\left(\left\|\bar{h}\right\|_{\tau-r}^{\tau-n^{-\nu}d}\right)^2\Bigg\}\\
&=\begin{cases}\left(\frac{\varepsilon}{2}\right)^{-2}d^{-\frac{1}{\nu}}\left\{m^2(\tau-)F'_-(\tau)+\|m\|^2_{U_r(\tau)\setminus\{\tau\}}\bar{L}^{\frac{\nu+1}{\nu}}\nu\left(F'_-(\tau)\right)^{-\frac{1}{\nu}}\right\}, & \text{if } \lambda=\frac{\nu+1}{2}\\ 0, & \text{if } \frac{1}{2}\le\lambda<\frac{\nu+1}{2\nu}\end{cases}.
\end{aligned}$$

where the expression in braces is a suitable positive constant C. □

Lemma 4.16 If $\|m\|<\infty$, then for each intervall $[\gamma,\delta]\subseteq T_F$ there exists some constant $C>0$ such that

$$\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\bar{L}_n(t)-\bar{H}(t)\right|}{\bar{F}(t)}\ge\varepsilon\right)\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|m\|^2\left\{2\bar{F}(\delta)^{-1}+C\right\}$$

for all $\varepsilon>0$.

Proof. Follow the proof of Lemma 4.12 and get

$$\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\bar{L}_n(t)-\bar{H}(t)\right|}{\bar{F}(t)}\ge\varepsilon\right) \le \mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\bar{I}_n(t)\right|}{\bar{F}(t)}\ge\frac{\varepsilon}{2}\right)+\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\bar{D}_n(t)-\bar{H}(t)\right|}{\bar{F}(t)}\ge\frac{\varepsilon}{2}\right). \quad (4.21)$$

For the first term we conclude

$$\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\bar{I}_n(t)\right|}{\bar{F}(t)}\ge\frac{\varepsilon}{2}\right)\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|m\|^2\,2\bar{F}(\delta)^{-1}.$$

To estimate the second term in (4.21) again choose $\eta\in(1/2,1)$ and use a quantile transformation to get

$$\frac{\left|\bar{D}_n(t)-\bar{H}(t)\right|}{\bar{F}(t)}\le\frac{\|m\|}{1-\eta}\sup_{t<x<\infty}\frac{|(F(x)-F_n(x))/F(x)|}{F(x)^{-\eta}}.$$

The martingale in the proof of Lemma 4.12 is now replaced by $F(x)-F_n(x)/F(x)$, which is a zero-mean càdlàg backwards-martingale on $[\gamma,\infty)$ (see for instance [Kou02, p. 68]) and

$$\mathbb{E}\left(\frac{F(x)-F_n(x)}{F(x)}\right)^2=n^{-1}\frac{\bar{F}(x)}{F(x)}.$$

Moreover, $F(x)^{-\eta}$ is positive, nonincreasing and right continuous on $[\gamma,\infty)$. We use continuity of $\mathbb{P}$ from below and Lemma 4.4 to get

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\gamma\le t\le\delta}\frac{\left|\bar{D}_n(t)-\bar{H}(t)\right|}{\bar{F}(t)}\ge\frac{\varepsilon}{2}\right)\\
&=\lim_{w\uparrow\infty}\mathbb{P}\left(\sup_{\gamma\le t\le w}\frac{|(F(x)-F_n(x))/F(x)|}{F(x)^{-\eta}}\ge\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)\\
&\le\left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2}\lim_{w\uparrow\infty}\left\{F(w)^{2\eta}n^{-1}\frac{1-F(w)}{F(w)}+\int\limits_{(\gamma,w]}F(x-)^{2\eta}\,\mathrm{d}\left(-n^{-1}\frac{1-F}{F}\right)\right\}.
\end{aligned}$$

Again, the Integration by parts theorem for Lebesgue-Stieltjes integrals, see [HS75, Theorem 21.67] and [SW86, page 868-869], implies for each $a < b$

$$
\begin{aligned}
&\frac{1-F(a)}{F(a)} - \frac{1-F(b)}{F(b)} \\
&= -\left(\int_{(a,b]} 1-F_- \,\mathrm{d}\left(\frac{1}{F}\right) + \int_{(a,b]} \frac{1}{F} \,\mathrm{d}(1-F) \right) \\
&= -\left(\int_{(a,b]} \frac{F_- -1}{FF_-} \,\mathrm{d}(F) + \int_{(a,b]} \frac{-1}{F} \,\mathrm{d}(F) \right) \\
&= \int_{(a,b]} \frac{1}{FF_-} \,\mathrm{d}F \,.
\end{aligned}
$$

Thus, the measure corresponding to $(F-1)/nF$ is a measure with F-density $1/nFF_-$ and we complete the proof with

$$
\begin{aligned}
&\left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2} \lim_{w\uparrow\infty} \left\{ F(w)^{2\eta} n^{-1} \frac{1-F(w)}{F(w)} + \int\limits_{(\gamma,w]} F(x-)^{2\eta} \,\mathrm{d}\left(-n^{-1}\frac{1-F}{F}\right) \right\} \\
&= \left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2} n^{-1} \lim_{w\uparrow\infty} \left\{ F(w)^{2\eta-1}(1-F(w)) + \int\limits_{(\gamma,w]} \frac{F(x-)^{2\eta-1}}{F(x)} F(\,\mathrm{d}x) \right\} \\
&\leq \left(\frac{\varepsilon}{2}\frac{1-\eta}{\|m\|}\right)^{-2} n^{-1} \lim_{w\uparrow\infty} \left\{ F(w)^{2\eta-1}(1-F(w)) + \frac{1}{2\eta-1}(F(w)-F(\gamma))^{2\eta-1} \right\} \\
&\leq n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2} \frac{\|m\|^2}{(1-\eta)^2(2\eta-1)} \,.
\end{aligned}
$$

□

4.6 Results for W_n

Lemma 4.17 If $\nu > 0$, $D > 0$ and if $r > 0$ such that **(A2)-(A3)** holds for $U_r(\tau)$ with $U_r(\tau) \subseteq T_F$ and $\|m\|_{U_r(\tau)\setminus\{\tau\}} < \infty$, then there exists a constant $C > 0$ such that

$$
\limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{d\leq|t|\leq n^\nu r} \left\{ n^{\frac{\nu+1}{2}} \left| W_n(\tau+n^{-\nu}t) - W_n(\tau) \right| \geq D|t|^{\frac{\nu+1}{2\nu}} \right\} \right) \leq C d^{-\frac{1}{\nu}}
$$

holds for all $d > 0$. Moreover,

$$\limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r} \left\{ n^{\lambda\nu} \left| W_n(\tau + n^{-\nu}t) - W_n(\tau) \right| \ge D|t|^\lambda \right\} \right) = 0$$

for all $d > 0$ and $1/2 \le \lambda < \nu+1/2\nu$.

Proof. In the following use $D > 0$ as a generic positive constant (cf. Remark 4.1). Decompose $Y_i - 1/2(\alpha+\beta) = (Y_i - m(X_i)) + (m(X_i) - 1/2(\alpha+\beta))$ and write

$$W_n(\tau + n^{-\nu}t) - W_n(\tau) = B_{\nu,n}(t) + V^{(1)}_{\nu,n}(t) - \frac{\alpha+\beta}{2} V^{(2)}_{\nu,n}(t)\,, \tag{4.22}$$

where

$$B_{\nu,n}(t) = n^{-1} \sum_{i=1}^{n} \left(\mathbb{1}_{X_i \le \tau + n^{-\nu}t} - \mathbb{1}_{X_i\le\tau} \right)(Y_i - m(X_i))\ , \tag{4.23}$$

$$V^{(1)}_{\nu,n}(t) = n^{-1} \sum_{i=1}^{n} \left(\left(\mathbb{1}_{X_i \le \tau + n^{-\nu}t} - \mathbb{1}_{X_i\le\tau} \right) m(X_i) - \mathbb{E}\left[\left(\mathbb{1}_{X_i \le \tau + n^{-\nu}t} - \mathbb{1}_{X_i\le\tau} \right) m(X_i) \right] \right)\ , \tag{4.24}$$

$$V^{(2)}_{\nu,n}(t) = n^{-1} \sum_{i=1}^{n} \left(\left(\mathbb{1}_{X_i \le \tau + n^{-\nu}t} - \mathbb{1}_{X_i\le\tau} \right) - \mathbb{E}\left(\mathbb{1}_{X_i \le \tau + n^{-\nu}t} - \mathbb{1}_{X_i\le\tau} \right) \right)\ . \tag{4.25}$$

Then decompose

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r} \left\{ n^{\lambda\nu} \left| W_n(\tau + n^{-\nu}t) - W_n(\tau) \right| \ge D|t|^\lambda \right\} \right) \\
&\le \mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r} \left\{ n^{\lambda\nu} \left| B_{\nu,n}(t) \right| \ge D|t|^\lambda \right\} \right) + \mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r} \left\{ n^{\lambda\nu} \left| V^{(1)}_{\nu,n}(t) \right| \ge D|t|^\lambda \right\} \right) \\
&\qquad + \mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r} \left\{ n^{\lambda\nu} \left| \frac{\alpha+\beta}{2} V^{(2)}_{\nu,n}(t) \right| \ge D|t|^\lambda \right\} \right)\ .
\end{aligned} \tag{4.26}$$

Consider the first term in (4.26) and recall the processes E_n and $\bar{E}_n$ introduced in (4.1) and (4.2), respectively. Then

$$\mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r}\left\{n^{\lambda\nu}\left|B_{\nu,n}(t)\right|\ge D|t|^\lambda\right\}\right)$$
$$\le \mathbb{P}\left(\sup_{d\le t\le n^\nu r}\frac{|E_n(\tau+n^{-\nu}t)-E_n(\tau)|}{(n^{-\nu}t)^\lambda}\ge D\right)+\mathbb{P}\left(\sup_{d\le -t\le n^\nu r}\frac{\left|\bar{E}_n(\tau+n^{-\nu}t)-\bar{E}_n(\tau)\right|}{(-n^{-\nu}t)^\lambda}\ge D\right).$$

Now have a look at the second term in (4.26) and recall the definition of the processes L_n and $\bar{L}_n$ in (4.3) and (4.4)

$$\mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r}\left\{n^{\lambda\nu}\left|V^{(1)}_{\nu,n}(t)\right|\ge D|t|^\lambda\right\}\right)$$
$$\le \mathbb{P}\left(\sup_{d\le t\le n^\nu r}\frac{|(L_n(\tau+n^{-\nu}t)-H(\tau+n^{-\nu}t))-(L_n(\tau)-H(\tau))|}{(n^{-\nu}t)^\lambda}\ge D\right)$$
$$+\,\mathbb{P}\left(\sup_{d\le -t\le n^\nu r}\frac{\left|\left(\bar{L}_n(\tau+n^{-\nu}t)-\bar{H}(\tau+n^{-\nu}t)\right)-\left(\bar{L}(\tau)-\bar{H}(\tau)\right)\right|}{(-n^{-\nu}t)^\lambda}\ge D\right).$$

Consider the last term in (4.26). For the case where $1/2(\alpha+\beta)=0$ there would be nothing to prove. So if $|1/2(\alpha+\beta)|>0$ and if we choose $m=1$ then $V^{(2)}_{\nu,n}(t)=V^{(1)}_{\nu,n}(t)$. An application of Corollaries 4.6, 4.8, 4.11 and 4.15 yields

$$\limsup_{n\to\infty}\mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r}\left\{n^{\frac{\nu+1}{2}}\left|W_n(\tau+n^{-\nu}t)-W_n(\tau)\right|\ge D|t|^{\frac{\nu+1}{2\nu}}\right\}\right)\le Cd^{-\frac{1}{\nu}}$$

and, in case $1/2\le\lambda<\nu+1/2\nu$,

$$\limsup_{n\to\infty}\mathbb{P}\left(\bigcup_{d\le|t|\le n^\nu r}\left\{n^{\lambda\nu}\left|W_n(\tau+n^{-\nu}t)-W_n(\tau)\right|\ge D|t|^\lambda\right\}\right)=0\,. \qquad \square$$

Lemma 4.18 If $\nu>0$, $D>0$, $\gamma\in\mathbb{R}$ and if $r>0$ such that **(A2)-(A3)** holds for $U_r(\tau)$ with $U_r(\tau)\subseteq T_F$ and $\|m\|_{U_r(\tau)\setminus\{\tau\}}<\infty$, then there exists a constant $C>0$ such that

$$\limsup_{n\to\infty}\mathbb{P}\left(\bigcup_{|t|\le d}\left\{n^{\frac{\nu+1}{2}}\left|W_n(\tau+n^{-\nu}t)-W_n(\tau)\right|\ge Dd^\gamma\right\}\right)\le Cd^{-2\gamma+1}$$

holds for all $d > 0$. Moreover,

$$\limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{|t|\le d} \left\{ n^{\lambda\nu} \left| W_n(\tau + n^{-\nu}t) - W_n(\tau) \right| \ge Dd^{\gamma} \right\} \right) = 0$$

for all $d > 0$ and $\lambda\nu < {}^{\nu+1}/_2$.

Proof. Let $D > 0$ be a generic constant (cf. Remark 4.1) and $d > 0$. We follow the procedure from the proof above and decompose W_n

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{|t|\le d} \left\{ n^{\lambda\nu} \left| W_n(\tau + n^{-\nu}t) - W_n(\tau) \right| \ge Dd^{\gamma} \right\} \right) \\
&\le \mathbb{P}\left(\bigcup_{|t|\le d} \left\{ n^{\lambda\nu} \left| B_{\nu,n}(t) \right| \ge Dd^{\gamma} \right\} \right) + \mathbb{P}\left(\bigcup_{|t|\le d} \left\{ n^{\lambda\nu} \left| V^{(1)}_{\nu,n}(t) \right| \ge Dd^{\gamma} \right\} \right) \\
&\qquad + \mathbb{P}\left(\bigcup_{|t|\le d} \left\{ n^{\lambda\nu} \left| \frac{\alpha+\beta}{2} V^{(2)}_{\nu,n}(t) \right| \ge Dd^{\gamma} \right\} \right) . \qquad (4.27)
\end{aligned}$$

Consider the first term of the right hand side of (4.27)

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{|t|\le d} \left\{ n^{\lambda\nu} \left| B_{\nu,n}(t) \right| \ge Dd^{\gamma} \right\} \right) \\
&\le \mathbb{P}\left(\sup_{0\le t\le d} \left| E_n(\tau + n^{-\nu}t) - E_n(\tau) \right| \ge Dd^{\gamma} n^{-\lambda\nu} \right) \\
&\qquad + \mathbb{P}\left(\sup_{0\le -t\le d} \left| \bar{E}_n(\tau + n^{-\nu}t) - \bar{E}_n(\tau) \right| \ge Dd^{\gamma} n^{-\lambda\nu} \right) .
\end{aligned}$$

Applying Lemmas 4.5 and 4.7, while choosing $q = \bar{q} = 1$ and using assumptions (A2)-(A3) and A.5, yields

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\sup_{0\le t\le d} \left| E_n(\tau + n^{-\nu}t) - E_n(\tau) \right| \ge Dd^{\gamma} n^{-\lambda\nu} \right) \\
&= \limsup_{n\to\infty} \mathbb{P}\left(\sup_{\tau\le s\le \tau+n^{-\nu}d} \left| E_n(s) - E_n(\tau) \right| \ge Dd^{\gamma} n^{-\lambda\nu} \right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \limsup_{n\to\infty} n^{2\lambda\nu-1} d^{-2\gamma} D \int\limits_{(\tau,\tau+n^{-\nu}d]} V(x)F(\mathrm{d}x) \\
&= \begin{cases} d^{-2\gamma+1} D\, V(\tau+)F'_+(\tau)\,, & \text{if } \lambda\nu = \frac{\nu+1}{2} \\ 0\,, & \text{if } \lambda\nu < \frac{\nu+1}{2} \end{cases},
\end{aligned}
$$

and

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\sup_{0\leq -t\leq d} \left|\bar{E}_n(\tau+n^{-\nu}t) - \bar{E}_n(\tau)\right| \geq Dd^{\gamma} n^{-\lambda\nu} \right) \\
&= \limsup_{n\to\infty} \mathbb{P}\left(\sup_{\tau-n^{-\nu}d\leq s\leq \tau} \left|\bar{E}_n(s) - \bar{E}_n(\tau)\right| \geq Dd^{\gamma} n^{-\lambda\nu} \right) \\
&\leq \limsup_{n\to\infty} n^{2\lambda\nu-1} d^{-2\gamma} D \int\limits_{(\tau-n^{-\nu}d,\tau]} V(x)F(\mathrm{d}x) \\
&= \begin{cases} d^{-2\gamma+1} D\, V(\tau-)F'_-(\tau)\,, & \text{if } \lambda\nu = \frac{\nu+1}{2} \\ 0\,, & \text{if } \lambda\nu < \frac{\nu+1}{2} \end{cases}.
\end{aligned}
$$

The second term in (4.27) can be estimated by

$$
\begin{aligned}
&\mathbb{P}\left(\bigcup_{|t|\leq d} \left\{ n^{\lambda\nu} \left| V^{(1)}_{\nu,n}(t) \right| \geq Dd^{\gamma} \right\} \right) \\
&\leq \mathbb{P}\left(\sup_{0\leq t\leq d} \left| \big(L_n(\tau+n^{-\nu}t) - H(\tau+n^{-\nu}t)\big) - (L_n(\tau) - H(\tau)) \right| \geq Dd^{\gamma} n^{-\lambda\nu} \right) \\
&\qquad + \mathbb{P}\left(\sup_{0\leq -t\leq d} \left| \big(\bar{L}_n(\tau+n^{-\nu}t) - \bar{H}(\tau+n^{-\nu}t)\big) - \big(\bar{L}(\tau) - \bar{H}(\tau)\big) \right| \geq Dd^{\gamma} n^{-\lambda\nu} \right) \\
&= \mathbb{P}\left(\sup_{\tau\leq s\leq \tau+n^{-\nu}d} |(L_n(s) - H(s)) - (L_n(\tau) - H(\tau))| \geq Dd^{\gamma} n^{-\lambda\nu} \right) \\
&\qquad\qquad + \mathbb{P}\left(\sup_{\tau-n^{-\nu}d\leq s\leq \tau} \left| \big(\bar{L}_n(s) - \bar{H}(s)\big) - \big(\bar{L}(\tau) - \bar{H}(\tau)\big) \right| \geq Dd^{\gamma} n^{-\lambda\nu} \right).
\end{aligned}
$$

Now use Lemmas 4.10 and 4.14, while choosing $q = \bar{q} = 1$. Furthermore, $\|m\|_{U_\varepsilon(\tau)} \bar{L} n^{-\nu} d$ gives an appropriate upper bound for $|H(s) - H(\tau)|$ and $|\bar{H}(s) - \bar{H}(\tau)|$ restricted to the interval $[\tau, \tau + n^{-\nu}d]$ and $[\tau - n^{-\nu}d, \tau]$, respectively. Then

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{|t|\le d}\left\{n^{\lambda\nu}\left|V^{(1)}_{\nu,n}(t)\right| \ge Dd^{\gamma}\right\}\right)\\
&\le \limsup_{n\to\infty} Dd^{-2\gamma}n^{2\lambda\nu-1}\left\{\int\limits_{(\tau,\tau+n^{-\nu}d]} m^2(t)F(\mathrm{d}t) + \frac{F(\tau+n^{-\nu}d)}{1-F(\tau+n^{-\nu}d)}\,\|m\|^2_{U_\varepsilon(\tau)}\,\bar{L}^2 n^{-2\nu}d^2\right\}\\
&+\limsup_{n\to\infty} Dd^{-2\gamma}n^{2\lambda\nu-1}\left\{\int\limits_{(\tau-n^{-\nu}d,\tau]} m^2(t)F(\mathrm{d}t) + \frac{1-F(\tau-n^{-\nu}d)}{F(\tau-n^{-\nu}d)}\,\|m\|^2_{U_\varepsilon(\tau)}\,\bar{L}^2 n^{-2\nu}d^2\right\}\\
&= \begin{cases} d^{-2\gamma+1}D\,\|m\|^2_{U_\varepsilon(\tau)}\,, & \text{if } \lambda\nu = \frac{\nu+1}{2}\\ 0\,, & \text{if } \lambda\nu < \frac{\nu+1}{2}\end{cases}.
\end{aligned}
$$

Again, observe that $V^{(2)}_{\nu,n}(t) = V^{(1)}_{\nu,n}(t)$ for the special case when $m = 1$ and (without loss of generality assume $|1/2(\alpha+\beta)| > 0$) we get

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{|t|\le d}\left\{n^{\lambda\nu}\left|\frac{\alpha+\beta}{2}V^{(2)}_{\nu,n}(t)\right| \ge Dd^{\gamma}\right\}\right)\\
&= \begin{cases} d^{-2\gamma+1}D\,, & \text{if } \lambda\nu = \frac{\nu+1}{2}\\ 0\,, & \text{if } \lambda\nu < \frac{\nu+1}{2}\end{cases}.
\end{aligned}
\tag{4.28}
$$

□

5 Consistency

The author of [Kos08, p. 271] proved consistency of $(\tau_n, \alpha_n, \beta_n)$ for the special case where the regression function is a decision tree itself. Thereby it is supposed that the split point τ lies in a known compact interval $T \subseteq T_F$. A further consistency proof for arbitrary regression functions and $\tau \in T \subseteq T_F$ is given in [RK10, Corollary 3.13]. In this chapter we will show how to dispense with the assumption that τ is in a known compactum $T \subseteq T_F$. To prepare for we start by studying the process H_n and proving an extension of Chang's theorem in Section 5.1. The main result of this chapter will be formulated in Section 5.2, where additional results necessary for the main proof are given. If not stated otherwise, we work in the notation of Section 2.1.

First we consider the estimator τ_n for the split point τ. Therefore, recall the characterization of τ in Lemma 2.1(iv) as a maximizing point of

$$M(t) = \begin{cases} \frac{H^2(t)}{F(t)} + \frac{\bar{H}^2(t)}{\bar{F}(t)} & \text{if } t \in T_F \\ (\mathbb{E}Y)^2 & \text{if } t \notin T_F \end{cases},$$

an those of τ_n in Lemma 2.2(i) as a maximizing point of

$$M_n(t) := \begin{cases} \frac{H_n^2(t)}{F_n(t)} + \frac{\bar{H}_n^2(t)}{\bar{F}_n(t)} & \text{if } t \in T_{F_n} \\ \bar{Y}_n^2 & \text{if } t \notin T_{F_n} \end{cases}.$$

In order to avoid confusing notation, in this section we adopt the convention ${}^0/_0 = 0$. Then for instance $H_n^2(t)/F_n(t) = 0$ for all $t < X_{1:n}$ and $\bar{H}_n^2(t)/\bar{F}_n(t) = 0$ for all $t \geq X_{n:n}$. It will turn out to be valuable to write

$$M_n(t) - M(t) = R_n(t) + \bar{R}_n(t)\,, \tag{5.1}$$

where

$$R_n(t) := \begin{cases} \frac{H_n^2(t)}{F_n(t)} - \frac{H^2(t)}{F(t)} & \text{if } F(t) > 0 \\ 0 & \text{else} \end{cases}$$

and

$$\bar{R}_n(t) := \begin{cases} \frac{\bar{H}_n^2(t)}{\bar{F}_n(t)} - \frac{\bar{H}^2(t)}{\bar{F}(t)} & \text{if } F(t) < 1 \\ 0 & \text{else} \end{cases} .$$

For an arbitrary real random variable with distribution Q and distribution function F we follow [Wit85, Definition 1.16] and introduce the quantile function (also named left continuous inverse) $F^{-1} : [0,1] \to \bar{\mathbb{R}}$ with

$$F^{-1}(y) := \begin{cases} \sup\{x \in \mathbb{R};\ F(x) = 0\} & y = 0 \\ \inf\{x \in \mathbb{R};\ F(x) \geq y\} & y \in (0,1) \\ \inf\{x \in \mathbb{R}; F(x) = 1\} & y = 1\ . \end{cases}$$

In the sequel, set

$$\lambda := F^{-1}(0) \quad \text{and} \quad \rho := F^{-1}(1).$$

As mentioned in [Wit85, p. 19], $[\lambda, \rho] = \bar{T}_F$ (the closure of T_F) is the smallest closed interval $I \subseteq \bar{\mathbb{R}}$ with $Q(I) = 1$. Furthermore, $F(\lambda-) = 0 \leq F(\lambda)$ and $F(\rho-) \leq 1 = F(\rho)$, compare [Wit85, Lemma 1.17 e)]. In the following we distinguish the cases where F is continuous in λ and ρ (which implies that $F(\lambda) = 0$ and $F(\rho-) = 1$), or F is discontinuous in one of these two points.

5.1 Extension of Chang's theorem

The following Lemma is a well-known result formulated in [SW86, p. 424].

Lemma 5.1 (Chang's Theorem, 1955) Let G_n, $n \in \mathbb{N}$, be the empirical distribution function corresponding to n, i. i. d. copies of a standard uniform distribution. If a sequence

$a_n \downarrow 0$ and $na_n \to \infty$, then

$$\sup_{a_n \le t \le 1} \left| \frac{G_n(t) - t}{t} \right| \xrightarrow{\mathbb{P}} 0 \,,$$

as $n \to \infty$.

Proof. The proof can be found in [Cha55]. □

The process H_n was also named *marked empirical process* and was examined in detail in [Stu97]. The next two lemmas give an extension of Chang's Theorem for the marked empirical process and an analogous result for $\bar{H}_n$.

Lemma 5.2 (Extension of Chang's Theorem) If $\|m\|, \|V\| < \infty$ and $(\lambda_n)_{n\in\mathbb{N}} \subseteq T_F$ is a sequence with $nF(\lambda_n) \to \infty$, then

$$\sup_{t \ge \lambda_n} \left| \frac{H_n(t) - H(t)}{F(t)} \right| \xrightarrow{\mathbb{P}} 0 \,,$$

as $n \to \infty$.

Proof. Fix $\varepsilon > 0$ and set $\rho = F^{-1}(1)$. Note that

$$\left\{ \sup_{t \ge \lambda_n} \left| \frac{H_n(t) - H(t)}{F(t)} \right| > \varepsilon \right\} \subseteq \left\{ \sup_{\lambda_n \le t < \rho} \left| \frac{H_n(t) - H(t)}{F(t)} \right| > \varepsilon \right\} \cup \left\{ \sup_{t \ge \rho} \left| \frac{H_n(t) - H(t)}{F(t)} \right| > \varepsilon \right\} . \tag{5.2}$$

Since we even have

$$\sup_{t \ge \rho} \frac{H_n(t) - H(t)}{F(t)} = \bar{Y}_n - \mathbb{E}Y \to 0 \quad \mathbb{P}-\text{a.s.}\,,$$

as $n \to \infty$, we only need to pay attention to the first set on the right hand side of (5.2). Now recall the defintion of the processes E_n and L_n in (4.1) and (4.3) and decompose

$$\begin{aligned} H_n(t) - H(t) =& \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}} (Y_i - m(X_i)) + \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}} m(X_i) - H(t) \\ =& E_n(t) + L_n(t) - H(t) \,. \end{aligned}$$

If we choose a sequence $(\rho_k)_{k\in\mathbb{N}} \subseteq T_F$ with $\rho_k \uparrow \rho$, then by continuity of $\mathbb{P}$ from below

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\lambda_n \le t < \rho}\left|\frac{H_n(t)-H(t)}{F(t)}\right| > \varepsilon\right) \\
&= \lim_{k\to\infty} \mathbb{P}\left(\sup_{\lambda_n \le t \le \rho_k}\left|\frac{H_n(t)-H(t)}{F(t)}\right| \ge \varepsilon\right) \\
&\le \lim_{k\to\infty} \mathbb{P}\left(\sup_{\lambda_n \le t \le \rho_k}\frac{|E_n(t)|}{F(t)} \ge \frac{\varepsilon}{2}\right) + \lim_{k\to\infty} \mathbb{P}\left(\sup_{\lambda_n \le t \le \rho_k}\frac{|L_n(t)-H(t)|}{F(t)} \ge \frac{\varepsilon}{2}\right). \qquad (5.3)
\end{aligned}$$

Applying Lemma 4.5, with $u = -\infty$ and $q = F$, and using a quantile transformation (compare the proof of Lemma 4.12) yields for the first term on the right hand side of (5.3)

$$\begin{aligned}
&\lim_{k\to\infty} \mathbb{P}\left(\sup_{\lambda_n \le t \le \rho_k}\frac{|E_n(t)|}{F(t)} \ge \frac{\varepsilon}{2}\right) \\
&\le \lim_{k\to\infty} n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\left[2F(\lambda_n)^{-2}\int_{(-\infty,\lambda_n]} V(x)\,F(\,\mathrm{d}x) + \int_{(\lambda_n,\rho_k]} F(x)^{-2}V(x)F(\,\mathrm{d}x)\right] \\
&\le \lim_{k\to\infty} n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|V\|\left[2F(\lambda_n)^{-1} + \left(F(\lambda_n)^{-1} - F(\rho_k)^{-1}\right)\right] \\
&\le n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2} 3\,\|V\|\,F(\lambda_n)^{-1}\,.
\end{aligned}$$

By Lemma 4.12 and $F(\rho_k) < 1$ for each $k \in \mathbb{N}$ we get for the second term in (5.3)

$$\lim_{k\to\infty} \mathbb{P}\left(\sup_{\lambda_n \le t \le \rho_k}\frac{|L_n(t)-H(t)|}{F(t)} \ge \frac{\varepsilon}{2}\right) \le n^{-1}\left(\frac{\varepsilon}{4}\right)^{-2}\|m\|^2\left\{2F(\lambda_n)^{-1} + C\right\}\,,$$

where $C > 0$ is some constant. Finally, with $nF(\lambda_n) \uparrow \infty$ then

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{t \ge \lambda_n}\left|\frac{H_n(t)-H(t)}{F(t)}\right| \ge \varepsilon\right) = 0\,. \qquad \square$$

Lemma 5.3 If $\|m\|, \|V\| < \infty$ and if $(\rho_n)_{n\in\mathbb{N}} \subseteq T_F$ is a sequence with $n(1-F(\rho_n-)) \uparrow \infty$, then

$$\sup_{t<\rho_n}\left|\frac{\bar{H}_n(t)-\bar{H}(t)}{\bar{F}(t)}\right| \xrightarrow{\mathbb{P}} 0\,,$$

as $n \to \infty$.

Proof. Set $\varepsilon > 0$ and $\lambda = F^{-1}(0)$. Decompose

$$\left\{\sup_{t<\rho_n}\left|\frac{\bar{H}_n(t)-\bar{H}(t)}{\bar{F}(t)}\right| > \varepsilon\right\} = \left\{\sup_{\lambda\le t<\rho_n}\left|\frac{\bar{H}_n(t)-\bar{H}(t)}{\bar{F}(t)}\right| > \varepsilon\right\} \cup \left\{\sup_{t<\lambda}\left|\frac{\bar{H}_n(t)-\bar{H}(t)}{\bar{F}(t)}\right| > \varepsilon\right\} . \tag{5.4}$$

Similar to the proof of Lemma 5.2 we only need to consider the first set on the right hand side of (5.4). With the definition of $\bar{E}_n$ und $\bar{L}_n$ in (4.1) and (4.3) then

$$\bar{H}_n(t)-\bar{H}(t) = \bar{E}_n(t) + \bar{L}_n(t) - \bar{H}(t) .$$

Let $(\lambda_k)_{k\in\mathbb{N}} \subseteq T_F$ be a sequence with $\lambda_k \downarrow \lambda$ and for each $n \in \mathbb{N}$ let $(\rho_{n,l})_{l\in\mathbb{N}} \subseteq T_F$ be a sequence with $\rho_{n,l} \uparrow \rho_n$. Now use right-continuity of the process and continuity from below of $\mathbb{P}$ to see that

$$\begin{aligned} &\mathbb{P}\left(\sup_{\lambda\le t<\rho_n}\left|\frac{\bar{H}_n(t)-\bar{H}(t)}{\bar{F}(t)}\right| > \varepsilon\right) \\ &\le \lim_{l\to\infty}\lim_{k\to\infty}\mathbb{P}\left(\sup_{\lambda_k\le t\le\rho_{n,l}}\frac{\left|\bar{E}_n(t)\right|}{\bar{F}(t)} \ge \frac{\varepsilon}{2}\right) + \lim_{l\to\infty}\lim_{k\to\infty}\mathbb{P}\left(\sup_{\lambda_k\le t\le\rho_{n,l}}\frac{\left|\bar{L}_n(t)-\bar{H}(t)\right|}{\bar{F}(t)} \ge \frac{\varepsilon}{2}\right) . \end{aligned} \tag{5.5}$$

By Lemma 4.7 and the same quantile transformation we obtain for the first term in (5.3)

$$\begin{aligned} &\lim_{l\to\infty}\lim_{k\to\infty}\mathbb{P}\left(\sup_{\lambda_k\le t\le\rho_{n,l}}\frac{\left|\bar{E}_n(t)\right|}{\bar{F}(t)} \ge \frac{\varepsilon}{2}\right) \\ &\le \lim_{l\to\infty}\lim_{k\to\infty} n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\left[2\bar{F}(\rho_{n,l})^{-2}\int\limits_{(\rho_{n,l},\infty)} V(x)F(\mathrm{d}x) + \int\limits_{(\lambda_k,\rho_{n,l}]}\bar{F}(x-)^{-2}V(x)F(\mathrm{d}x)\right] \\ &\le \lim_{l\to\infty}\lim_{k\to\infty} n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|V\|\left[2\bar{F}(\rho_{n,l})^{-1} + \left(\bar{F}(\rho_{n,l})^{-1} - \bar{F}(\lambda_k)^{-1}\right)\right] \\ &\le \lim_{l\to\infty} n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|V\|\,3\bar{F}(\rho_{n,l})^{-1} \\ &= n^{-1}\left(\frac{\varepsilon}{2}\right)^{-2}\|V\|\,3\bar{F}(\rho_n-)^{-1} . \end{aligned}$$

Continue with the second term in (5.5). By Lemma 4.16 and $F(\lambda_k) > 0$ for all $k \in \mathbb{N}$

there is a constant $C > 0$ such that

$$\lim_{l\to\infty}\lim_{k\to\infty}\mathbb{P}\left(\sup_{\lambda_k\le t\le\rho_{n,l}}\frac{\left|\bar{L}_n(t)-\bar{H}(t)\right|}{\bar{F}(t)}\ge\frac{\varepsilon}{2}\right)$$
$$\le\lim_{l\to\infty} n^{-1}\left(\frac{\varepsilon}{4}\right)^{-2}\|m\|^2\left\{2\bar{F}(\rho_{n,l})^{-1}+C\right\}$$
$$=n^{-1}\left(\frac{\varepsilon}{4}\right)^{-2}\|m\|^2\left\{2\bar{F}(\rho_n-)^{-1}+C\right\}\ .$$

Finally, with $n(1-F(\rho_n-))\uparrow\infty$ then

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{t<\rho_n}\left|\frac{\bar{H}_n(t)-\bar{H}(t)}{\bar{F}(t)}\right|\ge\varepsilon\right)=0\ .$$ □

5.2 Consistency of the estimator

If the distribution function F is continuous on the boundary of the set T_F, the requirement of τ lying within a known compactum $T\subseteq T_F$ may be disregarded. The following results therefore strongly depends on the shape of F on the boundary of T_F.

Theorem 5.4 Suppose that assumption (A1) holds with

$$\sup_{t\in\mathbb{R}} M(t) > \sup\{M(t);\ |t-\tau|>\varepsilon\}\ . \tag{5.6}$$

(i) If $F(F^{-1}(0))>0$ and $F(F^{-1}(1)-)<1$ then

$$\tau_n\longrightarrow\tau\quad\mathbb{P}\text{-almost surely}\,,$$

as $n\to\infty$. If in addition F is continuous in an ε-neighborhood $U_\varepsilon(\tau)$ of τ, then

$$(\tau_n,\alpha_n,\beta_n)\longrightarrow(\tau,\alpha,\beta)\quad\mathbb{P}\text{-almost surely}\,,$$

as $n\to\infty$.

(ii) If otherwise $F(F^{-1}(0))=0$ or $F(F^{-1}(1)-)=1$ and if $\|m\|\,,\|V\|<\infty$ then

$$\tau_n\xrightarrow{\mathbb{P}}\tau\ ,$$

as $n \to \infty$. If in addition F is continuous in an ε-neighborhood $U_\varepsilon(\tau)$ of τ, then

$$(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)\,,$$

as $n \to \infty$.

Condition (5.6) for τ is often called 'well-separated'. For a continuous distribution function F it can be derived immediately as follows.

Corollary 5.5 Suppose that assumption (A1) holds, F is continuous and $\|m\|, \|V\| < \infty$, then

$$(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)\,,$$

as $n \to \infty$.

Proof. Since F is continuous, M is as well and $F(F^{-1}(0)) = 0$ and $F(F^{-1}(1)-) = 1$. Furthermore, τ is the unique maximizer of M (cf. Lemma 2.1(iv)) and, hence, condition (5.6) follows from [Kos08, Lemma 14.3]. The claim then follows from Theorem 5.4. □

We first present some preliminaries and analyze the processes R_n and $\bar{R}_n$. At the end of this section we will prove Theorem 5.4

Lemma 5.6 If $F(F^{-1}(0)) > 0$, then

$$\sup_{t\in\mathbb{R}} |R_n(t)| \longrightarrow 0 \quad \mathbb{P}\text{-almost surely}\,,$$

as $n \to \infty$, and if $F(F^{-1}(1)-) < 1$, then

$$\sup_{t\in\mathbb{R}} \left|\bar{R}_n(t)\right| \longrightarrow 0 \quad \mathbb{P}\text{-almost surely}\,,$$

as $n \to \infty$.

Proof. Set $\lambda = F^{-1}(0)$. Then $F(t) > 0$ for all $t \geq \lambda$ and we can rewrite

$$R_n(t) = \frac{H_n^2(t)}{F_n(t)} - \frac{H^2(t)}{F(t)} = [H_n(t) - H(t)]\left[\frac{H_n(t)}{F_n(t)} + \frac{H(t)}{F(t)}\right] + [F(t) - F_n(t)]\left[\frac{H_n(t)}{F_n(t)}\frac{H(t)}{F(t)}\right]$$

for all $t \geq \lambda$. Note that $R_n(t) = 0$ if $t < \lambda$. Since $F(\lambda) > 0$, we follow that

$$\sum_{n\in\mathbb{N}} \mathbb{P}(\{X_1 > \lambda\} \cap \ldots \cap \{X_n > \lambda\}) = \sum_{n\in\mathbb{N}} (1 - F(\lambda))^n = F(\lambda)^{-1} < \infty$$

and the Borel-Cantelli lemma ([Kle06, Proposition 2.7]) implies that with probability one there exists some $N \in \mathbb{N}$ such that $F_n(\lambda) > 0$ for all $n \geq N$. Thus for sufficiently large $n \in \mathbb{N}$

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} |R_n(t)| \\
&\leq \sup_{t\geq\lambda} |R_n(t)| \\
&\leq \|H_n - H\| \sup_{t\geq\lambda}\left[\frac{|H_n(t)|}{F_n(t)} + \frac{|H(t)|}{F(t)}\right] + \|F_n - F\| \sup_{t\geq\lambda}\left[\frac{|H_n(t)|}{F_n(t)}\frac{|H(t)|}{F(t)}\right] \\
&\leq \|H_n - H\| \left[\frac{\frac{1}{n}\sum_{i=1}^n |Y_i|}{F_n(\lambda)} + \frac{\mathbb{E}|Y|}{F(\lambda)}\right] + \|F_n - F\| \left[\frac{\frac{1}{n}\sum_{i=1}^n |Y_i|}{F_n(\lambda)}\frac{\mathbb{E}|Y|}{F(\lambda)}\right].
\end{aligned}$$

By the Strong Law of Large Numbers (SLLN) $n^{-1}\sum_{i=1}^n |Y_i| \to \mathbb{E}|Y|$ almost surely and $F_n(\lambda) \to F(\lambda)$ almost surely as $n \to \infty$. Due to the Glivenko-Cantelli theorem (cf. [SW86, Theorem 3.3]) and $\|H_n - H\| \to 0$ almost surely ([Stu97, p. 615] and [Til07, p. 18-21] [1]) we have $\sup_{t\in\mathbb{R}} |R_n(t)| \to 0$ almost surely. In a similar way rearrange $\bar{R}_n(t)$ and estimate

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} \left|\bar{R}_n(t)\right| \\
&\leq \left\|\bar{H}_n - \bar{H}\right\| \sup_{t<\rho}\left[\frac{\left|\bar{H}_n(t)\right|}{\bar{F}_n(t)} + \frac{\left|\bar{H}(t)\right|}{\bar{F}(t)}\right] + \|F_n - F\| \sup_{t<\rho}\left[\frac{\left|\bar{H}_n(t)\right|}{\bar{F}_n(t)}\frac{\left|\bar{H}(t)\right|}{\bar{F}(t)}\right] \\
&\leq \left\|\bar{H}_n - \bar{H}\right\| \left[\frac{\frac{1}{n}\sum_{i=1}^n |Y_i|}{\bar{F}_n(\rho-)} + \frac{\mathbb{E}|Y|}{\bar{F}(\rho-)}\right] + \|F_n - F\| \left[\frac{\frac{1}{n}\sum_{i=1}^n |Y_i|}{\bar{F}_n(\rho-)}\frac{\mathbb{E}|Y|}{\bar{F}(\rho-)}\right].
\end{aligned}$$

Note that $\|\bar{H}_n - \bar{H}\| \leq |\bar{Y}_n - \mathbb{E}Y| + \|H_n - H\|$. With the same arguments as above finally $\sup_{t\in\mathbb{R}} |\bar{R}_n(t)| \to 0$ almost surely. □

Lemma 5.7 If $\|m\| < \infty$ then for each sequence $(\lambda_n)_{n\in\mathbb{N}} \subseteq T_F$ with $F(\lambda_n) \downarrow 0$ one has

$$\sup_{t<\lambda_n} |R_n(t)| \xrightarrow{\mathbb{P}} 0\,,$$

as $n \to \infty$.

[1] was stated in [Stu97, p. 615] without proof and proved in [Til07, p. 18-21]

Proof. By the Cauchy-Schwarz inequality one has

$$H_n^2(t) = \frac{1}{n^2}\left(\sum_{i=1}^n \mathbb{1}_{X_i \le t} Y_i\right)^2 \le \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i \le t} Y_i^2 \, F_n(t)\,. \tag{5.7}$$

Analogously, with Cauchy-Schwarz

$$H^2(t) \overset{(2.3)}{=} \left(\mathbb{E}\left(\mathbb{1}_{X \le t} m(X)\right)\right)^2 \le \mathbb{E}\left(\mathbb{1}_{X \le t}\right)\mathbb{E}\left(\mathbb{1}_{X \le t} m^2(X)\right) = \|m\|^2\, F^2(t)$$

and, therefore,

$$\sup_{t<\lambda_n} |R_n(t)| \le \sup_{t<\lambda_n}\left[\frac{H_n^2(t)}{F_n(t)} + \frac{H^2(t)}{F(t)}\right] \le \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i<\lambda_n\}} Y_i^2 + \|m\|^2\, F(\lambda_n)\,.$$

Set $\varepsilon > 0$ and use Markov's inequality to get

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i<\lambda_n\}} Y_i^2 \ge \varepsilon\right) \le \varepsilon^{-1}\mathbb{E}\left(\mathbb{1}_{\{X<\lambda_n\}} E\left(Y^2|X\right)\right) = \varepsilon^{-1} \int\limits_{(-\infty,\lambda_n)} K(x) F(\mathrm{d}x)\,,$$

where $K(x) := \mathbb{E}(Y^2|X = x)$. Since

$$\int K(x)F(\mathrm{d}x) = \mathbb{E}\left(\mathbb{E}\left(Y^2|X\right)\right) = \mathbb{E}Y^2 < \infty\,,$$

and $F(\lambda_n) \downarrow 0$, an application of Lebesgue's dominated convergence theorem with the dominating function $K(x)$ finally yields

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{t<\lambda_n} |R_n(t)| \ge \varepsilon\right) = 0\,. \qquad \square$$

Lemma 5.8 If $\|m\|, \|V\| < \infty$, then for each sequence $(\lambda_n)_{n\in\mathbb{N}} \subseteq T_F$ with $F(\lambda_n) \downarrow 0$ and $nF(\lambda_n) \uparrow \infty$, one has

$$\sup_{t\ge\lambda_n} |R_n(t)| \xrightarrow{\mathbb{P}} 0\,,$$

as $n \to \infty$.

Proof. First rewrite R_n for all $t \in \mathbb{R}$ with $F(t) > 0$ as

$$\begin{aligned}
R_n(t) &= \frac{H_n^2(t)}{F(t)} - \frac{H^2(t)}{F(t)} - \frac{H_n^2(t)}{F_n(t)} \frac{F_n(t) - F(t)}{F(t)} \\
&= \left[\frac{H_n(t) - H(t)}{F(t)} (H_n(t) + H(t)) \right] - \left[\frac{H_n^2(t)}{F_n(t)} \frac{F_n(t) - F(t)}{F(t)} \right] \\
&:= R_n^{(1)}(t) - R_n^{(2)}(t) \,. \qquad (5.8)
\end{aligned}$$

For each $\varepsilon > 0$ then

$$\mathbb{P}\left(\sup_{t \geq \lambda_n} |R_n(t)| \geq \varepsilon \right) \leq \mathbb{P}\left(\sup_{t \geq \lambda_n} \left| R_n^{(1)}(t) \right| \geq \frac{\varepsilon}{2} \right) + \mathbb{P}\left(\sup_{t \geq \lambda_n} \left| R_n^{(2)}(t) \right| \geq \frac{\varepsilon}{2} \right) .$$

Set $\rho = F^{-1}(1)$ and note that $R_n^{(2)}(t) = 0$ for all $t > \rho$ almost surely. Let $U_1, \dots, U_n$ be i. i. d. copies of a standard uniform distribution with empirical distribution function G_n, then $F^{-1}(U_i)$ for $i = 1, \dots, n$ are i. i. d. copies of F and $G_n \circ F \stackrel{\mathscr{L}}{=} F_n$ (see [Wit85, Lemma 2.29] and [WM95, p. 551]). By the Cauchy-Schwarz inequality, compare with (5.7), then

$$\begin{aligned}
\sup_{t \geq \lambda_n} \left| R_n^{(2)}(t) \right| &= \sup_{\lambda_n \leq t \leq \rho} \left| \frac{H_n^2(t)}{F_n(t)} \frac{F_n(t) - F(t)}{F(t)} \right| \\
&\leq n^{-1} \sum_{i=1}^n Y_i^2 \sup_{\lambda_n \leq t \leq \rho} \left| \frac{F_n(t) - F(t)}{F(t)} \right| \\
&\leq n^{-1} \sum_{i=1}^n Y_i^2 \sup_{F(\lambda_n) \leq x \leq 1} \left| \frac{G_n(x) - x}{x} \right| .
\end{aligned}$$

Since $F(\lambda_n) \downarrow 0$ and $nF(\lambda_n) \uparrow \infty$ we can use Chang's theorem (Lemma 5.1) and the fact that $n^{-1} \sum_{i=1}^n Y_i^2 \to \mathbb{E}Y^2$ almost surely as $n \to \infty$ and conclude

$$\lim_{n \to \infty} \mathbb{P}\left(\sup_{t \geq \lambda_n} \left| R_n^{(2)}(t) \right| \geq \frac{\varepsilon}{2} \right) = 0 \,.$$

Furthermore,

$$\sup_{t \geq \lambda_n} \left| R_n^{(1)}(t) \right| \leq \left(n^{-1} \sum_{i=1}^n |Y_i| + \mathbb{E}\,|Y| \right) \sup_{t \geq \lambda_n} \left| \frac{H_n(t) - H(t)}{F(t)} \right| .$$

By the SLLN the first term converges almost surely to some $0 \leq c < \infty$, as $n \to \infty$. For

$F(\lambda_n) \downarrow 0$ and $nF(\lambda_n) \uparrow \infty$, by Lemma 5.2 we finally see that

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{t\geq\lambda_n} \left|R_n^{(1)}(t)\right| \geq \frac{\varepsilon}{2} \right) = 0 \,. \qquad \square$$

To be complete, we add a corresponding result for the process $\bar{R}_n$.

Lemma 5.9 If $\|m\| < \infty$ then for each sequence $(\rho_n)_{n\in\mathbb{N}} \subseteq T_F$ with $F(\rho_n) \uparrow 1$ one has

$$\sup_{t\geq\rho_n} \left|\bar{R}_n(t)\right| \xrightarrow{\mathbb{P}} 0 \,,$$

as $n \to \infty$.

Proof. Use the same arguments used to proof Lemma 5.7. $\square$

Lemma 5.10 If $\|m\|, \|V\| < \infty$, then for each sequence $(\rho_n)_{n\in\mathbb{N}} \subseteq T_F$ with $F(\rho_n-) \uparrow 1$ and $n(1-F(\rho_n-)) \uparrow \infty$

$$\sup_{t<\rho_n} \left|\bar{R}_n(t)\right| \xrightarrow{\mathbb{P}} 0 \,,$$

as $n \to \infty$.

Proof. Similar to the proof of Lemma 5.8 rewrite $\bar{R}_n$ for all $t \in \mathbb{R}$ with $\bar{F}(t) > 0$as

$$\bar{R}_n(t) = \left[\frac{\bar{H}_n(t) - \bar{H}(t)}{\bar{F}(t)} \left(\bar{H}_n(t) + \bar{H}(t)\right) \right] - \left[\frac{\bar{H}_n^2(t)}{\bar{F}_n(t)} \frac{\bar{F}_n(t) - \bar{F}(t)}{\bar{F}(t)} \right] := \bar{R}_n^{(1)}(t) + \bar{R}_n^{(2)}(t) \,. \quad (5.9)$$

Fix $\varepsilon > 0$ and use (5.9) for the estimate

$$\mathbb{P}\left(\sup_{t<\rho_n} \left|\bar{R}_n(t)\right| \geq \varepsilon \right) \leq \mathbb{P}\left(\sup_{t<\rho_n} \left|\bar{R}_n^{(1)}(t)\right| \geq \frac{\varepsilon}{2} \right) + \mathbb{P}\left(\sup_{t<\rho_n} \left|\bar{R}_n^{(2)}(t)\right| \geq \frac{\varepsilon}{2} \right) . \qquad (5.10)$$

Since for a standard uniform random variable $U \stackrel{\mathscr{L}}{=} 1-U$ we get the following identity (in the sense of identically distributed) for the corresponding càdlàg empirical distribution function G_n

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq t\}} = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{1-U_i < 1-t\}} \stackrel{\mathscr{L}}{=} 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{U_i < 1-t\}} = 1 - G_n((1-t)-)$$

and thus $\bar{F}_n \stackrel{\mathscr{L}}{=} 1 - G_n \circ F \stackrel{\mathscr{L}}{=} G_n(\bar{F}-)$ (compare [Wit85, Lemma 2.29] and [WM95, p. 551]). Set $\lambda = F^{-1}(0)$ and note that $\bar{R}_n^{(2)}(t) = 0$ almost surely for all $t < \lambda$. Together with A.7

this implies

$$\begin{aligned}
\sup_{t<\rho_n}\left|\bar{R}_n^{(2)}(t)\right| &= \sup_{\lambda\le t<\rho_n}\left|\frac{\bar{H}_n^2(t)}{\bar{F}_n(t)}\frac{\bar{F}_n(t)-\bar{F}(t)}{\bar{F}(t)}\right| \\
&\le n^{-1}\sum_{i=1}^n Y_i^2 \sup_{\lambda\le t<\rho_n}\left|\frac{\bar{F}_n(t)-\bar{F}(t)}{\bar{F}(t)}\right| \\
&= n^{-1}\sum_{i=1}^n Y_i^2 \sup_{\lambda\le t<\rho_n}\left|\frac{G_n(\bar{F}(t)-)-\bar{F}(t)}{\bar{F}(t)}\right| \\
&= n^{-1}\sum_{i=1}^n Y_i^2 \sup_{\bar{F}(\rho_n-)<x\le\bar{F}(\lambda)}\left|\frac{G_n(x-)-x}{x}\right| \\
&= n^{-1}\sum_{i=1}^n Y_i^2 \sup_{\bar{F}(\rho_n-)\le x<\bar{F}(\lambda)}\left|\frac{G_n(x)-x}{x}\right| .
\end{aligned}$$

Similar to the proof of Lemma 5.8, for $\bar{F}(\rho_n-)\downarrow 0$ and $n(1-F(\rho_n-))\uparrow\infty$, Lemma 5.1 implies

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{t<\rho_n}\left|\bar{R}_n^{(2)}(t)\right|\ge\frac{\varepsilon}{2}\right)=0 .$$

Moreover,

$$\sup_{t<\rho_n}\left|\bar{R}_n^{(1)}(t)\right| \le \left(n^{-1}\sum_{i=1}^n|Y_i|+\mathbb{E}|Y|\right)\sup_{t<\rho_n}\left|\frac{\bar{H}_n(t)-\bar{H}(t)}{\bar{F}(t)}\right|$$

and by the SLLN and Lemma 5.3 we finally get

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{t<\rho_n}\left|\bar{R}_n^{(1)}(t)\right|\ge\frac{\varepsilon}{2}\right)=0 . \qquad \square$$

Now we show under which assumptions on the distribution function F the existence of sequences $(\lambda_n)_{n\in\mathbb{N}}\subseteq T_F$ and $(\rho_n)_{n\in\mathbb{N}}\subseteq T_F$ with $F(\lambda_n)\downarrow 0$, $F(\rho_n-)\uparrow 1$, $nF(\lambda_n)\to\infty$ and $n(1-F(\rho_n-))\to\infty$ is ensured.

Lemma 5.11 Suppose that F is a distribution function with $F(F^{-1}(0))=0$ and $F(F^{-1}(1)-)=1$. If $(a_n)_{n\in\mathbb{N}}\subseteq(0,1)$ is a sequence such that $a_n\downarrow 0$ and $na_n\to\infty$, then

(i) $F^{-1}(a_n)\in T_F$ for each $n\in\mathbb{N}$ and $F(F^{-1}(a_n))\downarrow 0$ and $nF(F^{-1}(a_n))\to\infty$,

(ii) $F^{-1}(1-a_n)\in T_F$ for each $n\in\mathbb{N}$ and $F(F^{-1}(1-a_n)-)\uparrow 1$ and $n(1-F(F^{-1}(1-a_n)-))\to\infty$.

Proof. (i) Settig $\lambda_n = F^{-1}(a_n)$, we can conclude from [Wit85, Lemma 1.17 e)] that $0 < a_n \le F(F^{-1}(a_n)) = F(\lambda_n)$. In order to show that $F(\lambda_n) < 1$, conversely, suppose that $F(\lambda_n) = 1$. By [Wit85, Lemma 1.17 a)] then $\lambda_n \ge F^{-1}(1)$. But, by Lemma [Wit85, Lemma 1.18 b)] and $a_n < 1$ this is equivalent to the condition that $F^{-1}(1)$ is a discontinuity point of F, a contradiction to $F(F^{-1}(1)-) = 1$. Hence we have $(\lambda_n)_{n\in\mathbb{N}} \subseteq T_F$. Furthermore, observe that $(\lambda_n)_{n\in\mathbb{N}}$ is monotonically non-increasing, since F^{-1} is monotonically non-decreasing (cf. [Wit85, Lemma 1.15]). The definition of $F^{-1}(0)$ was chosen such that $\bar{T}_F = [F^{-1}(0), F^{-1}(1)]$. Therefore, $\lambda_n \downarrow F^{-1}(0)$ and, by the right-continuity of F and $F(F^{-1}(0)) = 0$, we have $F(\lambda_n) \downarrow 0$. Finally, $nF(\lambda_n) \ge na_n$ (cf. [Wit85, Lemma 1.17 e)]) and, hence, $nF(\lambda_n) \to \infty$.

(ii) Set $\rho_n = F^{-1}(1 - a_n)$. With arguments similar to those in (i) show that $(\rho_n)_{n\in\mathbb{N}} \subseteq T_F$ and that $(\rho_n)_{n\in\mathbb{N}}$ is monotonically non-decreasing with $\rho_n \uparrow F^{-1}(1)$. By the left-continuity of F_- and $F(F^{-1}(1)-) = 1$ we find that $F(\rho_n-) \uparrow 1$. Lastly, $F(\rho_n-) \le 1 - a_n$ (cf. [Wit85, Lemma 1.17 e)]) and thus $n(1 - F(\rho_n-)) \to \infty$. □

Corollary 5.12 (i) Suppose that $F(F^{-1}(0)) > 0$ and $F(F^{-1}(1)-) < 1$, then

$$\sup_{t\in\mathbb{R}} |M_n(t) - M(t)| \longrightarrow 0 \quad \mathbb{P}\text{-almost surely}$$

as $n \to \infty$.

(ii) If otherwise $F(F^{-1}(0)) = 0$ or $F(F^{-1}(1)-) = 1$ and if $\|m\|, \|V\| < \infty$, then

$$\sup_{t\in\mathbb{R}} |M_n(t) - M(t)| \xrightarrow{\mathbb{P}} 0\,,$$

as $n \to \infty$.

Proof. (i) Recall the decomposition from (5.1) to see that

$$\sup_{t\in\mathbb{R}} |M_n(t) - M(t)| \le \sup_{t\in\mathbb{R}} |R_n(t)| + \sup_{t\in\mathbb{R}} \left|\bar{R}_n(t)\right| \,.$$

If $F(F^{-1}(0)) > 0$ and $F(F^{-1}(1)-) < 1$, an application of Lemma 5.6 ensures the almost sure convergence.

(ii) Now assume that both $F(F^{-1}(0)) = 0$ and $F(F^{-1}(1)-) = 1$. Let $(a_n)_{n\in\mathbb{N}} \subseteq (0,1)$ be a sequence such that $a_n \downarrow 0$ and $na_n \to \infty$. For each $n \in \mathbb{N}$ set $\lambda_n := F^{-1}(a_n)$ and $\rho_n := F^{-1}(1 - a_n)$. By Lemma 5.11 then $(\lambda_n)_{n\in\mathbb{N}}, (\rho_n)_{n\in\mathbb{N}} \subseteq T_F$ with $F(\lambda_n) \downarrow 0$, $F(\rho_n-) \uparrow 1$,

$nF(\lambda_n) \to \infty$, and $n(1-F(\rho_n-)) \to \infty$. The decomposition of $M_n - M$ from Equation (5.1) implies that for all $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in\mathbb{R}} |M_n(t)-M(t)| \geq \varepsilon\right) \\
&\leq \mathbb{P}\left(\sup_{t<\lambda_n} |R_n(t)| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\sup_{t\geq\lambda_n} |R_n(t)| \geq \frac{\varepsilon}{2}\right) \\
&\qquad + \mathbb{P}\left(\sup_{t<\rho_n} \left|\bar{R}_n(t)\right| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\sup_{t\geq\rho_n} \left|\bar{R}_n(t)\right| \geq \frac{\varepsilon}{2}\right).
\end{aligned}$$

Applying Lemmas 5.7, 5.8, 5.9 and 5.10 then yields $\|M_n(t)-M(t)\| \xrightarrow{\mathbb{P}} 0$, as $n \to \infty$. Since almost sure convergence implies convergence in probability it follows for the 'mixed' cases (that means the cases $[F(F^{-1}(0)) > 0$ and $F(F^{-1}(1)-) = 1\ ]$ and $[F(F^{-1}(0)) = 0$ and $F(F^{-1}(1)-) < 1\ ]$) that $\|M_n(t)-M(t)\| \xrightarrow{\mathbb{P}} 0$, as $n \to \infty$. □

Proof of Theorem 5.4. (i) By Lemma 2.1(iv) and Remark 3.10, $\tau \in \operatorname{Argsup}(M)$ and by Lemma 2.2(i) and Remark 3.10, $\tau_n \in \operatorname{Argsup}(M_n)$ for each $n \in \mathbb{N}$. Since by Corollary 5.12(i), $\|M_n - M\| \to 0$ almost surely as $n \to \infty$, we can use Proposition 3.11 to get $\tau_n \to \tau$, almost surely as $n \to \infty$. If, in addition, F is continuous in $U_\varepsilon(\tau)$, then it is easy to check that τ is an interior point of T_F. Thus, there exists $\delta > 0$ such that $[\tau-\delta, \tau+\delta] \subseteq T_F$. Since $\tau_n \to \tau$, almost surely, then $\tau_n \in [\tau-\delta, \tau+\delta]$ and $F_n(\tau-\delta) > 0$ for sufficiently large $n \in \mathbb{N}$. Furthermore,

$$\begin{aligned}
|a_n(\tau_n) - a(\tau_n)| &\leq \sup_{t\in[\tau-\delta,\tau+\delta]} |a_n(t)-a(t)| \\
&= \sup_{t\in[\tau-\delta,\tau+\delta]} \left|\frac{H_n(t)}{F_n(t)} + \frac{H(t)}{F(t)}\right| \\
&= \sup_{t\in[\tau-\delta,\tau+\delta]} \left|\frac{H_n(t)-H(t)}{F_n(t)} + \frac{H(t)(F(t)-F_n(t))}{F_n(t)F(t)}\right| \\
&\leq \frac{\|H_n - H\|}{F_n(\tau-\delta)} + \frac{\mathbb{E}|Y|\ \|F_n - F\|}{F_n(\tau-\delta)F(\tau-\delta)},
\end{aligned}$$

which converges almost surely to zero by the Glivenko-Cantelli theorem and since $\|H_n - H\| \to 0$ almost surely ([Stu97, p. 615],[Til07, p. 18-21]). The same reasoning applied to $|\beta_n - \beta|$ allows to conclude $\beta_n \to \beta$ almost surely, and therefore $(\tau_n, \alpha_n, \beta_n) \to (\tau, \alpha, \beta)$ almost surely.

(ii) Following the same reasoning at the end of the proof of Corollary 5.12 it suffices to study the case where both $F(F^{-1}(0)) = 0$ and $F(F^{-1}(1-)) = 1$. We use Corollary 5.12(ii) to get $\|M_n - M\| \xrightarrow{\mathbb{P}} 0$ and apply Proposition 3.11 to conclude $\tau_n \xrightarrow{\mathbb{P}} \tau$, as $n \to \infty$. For each $n \in \mathbb{N}$ we have $\alpha_n = a_n(\tau_n)$, see Lemma 2.2(ii). Now let $(\alpha_{n'})_{n' \in \mathbb{N}}$, with $\alpha_{n'} = a_{n'}(\tau_{n'})$, be a subsequence of $(\alpha_n)_{n \in \mathbb{N}}$. Since $F(F^{-1}(0)) = 0$ we find that $\tau > F^{-1}(0)$. By the subsequence criterion for convergence in probability there exists $\delta > 0$ and some further sub-subsequence $(\tau_{n''})_{n'' \in \mathbb{N}} \subseteq U_\delta(\tau) \subseteq T_F$ such that $\tau_{n''} \to \tau$, almost surely, and $(\alpha_{n''})_{n'' \in \mathbb{N}}$ with $\alpha_{n''} = a_{n''}(\tau_{n''})$ is a subsequence of $(\alpha_{n'})_{n' \in \mathbb{N}}$. Then

$$|\alpha_{n''} - \alpha| \leq |a_{n''}(\tau_{n''}) - a(\tau_{n''})| + |a(\tau_{n''}) - a(\tau)| \ , \tag{5.11}$$

where $a(t)$ is the function from Lemma 2.1(ii). For sufficiently large $n'' \in \mathbb{N}$ we find

$$\begin{aligned}
|a_{n''}(\tau_{n''}) - a(\tau_{n''})| &\leq \sup_{t \in U_\delta(\tau)} |a_{n''}(t) - a(t)| \\
&= \sup_{t \in U_\delta(\tau)} \left| \frac{H_{n''}(t)}{F_{n''}(t)} - \frac{H(t)}{F(t)} \right| \\
&\leq \sup_{t \in U_\delta(\tau)} \left| \frac{H_{n''}(t) - H(t)}{F_{n''}(t)} + \frac{H(t)(F(t) - F_{n''}(t))}{F_{n''}(t)F(t)} \right| \\
&\leq \frac{\|H_{n''} - H\|}{F_{n''}(\tau - \delta)} + \frac{\mathbb{E}|Y| \ \|F_{n''} - F\|}{F_{n''}(\tau - \delta)F(\tau - \delta)} \ ,
\end{aligned}$$

which converges to zero almost surely, as $n'' \to \infty$, by the same arguments as above. Moreover, continuity of F in $U_\varepsilon(\tau)$ implies the continuity of $a(t)$ and, thus, the second term on the right hand side of (5.11) converges to zero almost surely as well. By a further application of the subsequence criterion then $\alpha_n \xrightarrow{\mathbb{P}} \alpha$. An analogous procedure allows to conclude $\beta_n \xrightarrow{\mathbb{P}} \beta$, and, finally, the hole vector $(\tau_n, \alpha_n, \beta_n)$ converges to (τ, α, β) in probability. □

6 Discontinuous Case

This chapter establishes the joint distributional convergence of the estimators in the case where the regression function is discontinuous at the split point. The key is to express the estimators multiplied with the corresponding convergence rates as an infimizer of a rescaled process and apply Proposition 3.12. We first state the discontinuity condition in Section 6.1. In Section 6.2 we introduce the rescaled process as a process in the multivariate Skorokhod space, identify its limit process and derive convergence in distribution. Section 6.3 deals with the condition of stochastic boundedness and Section 6.4 assembles all results and presents the main statement.

6.1 Regression function with a jump at τ

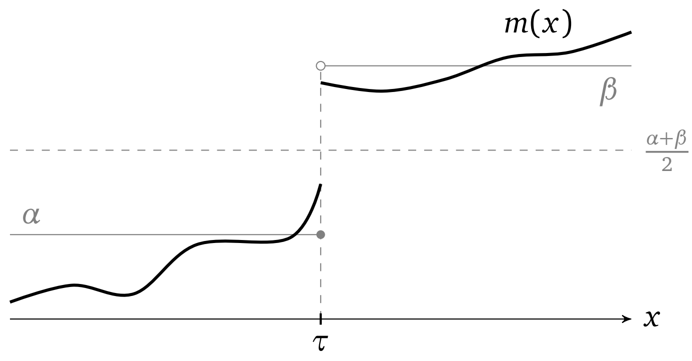
We give an exact formulation of the discontinuity condition.

(B1) The regression function m is of the form

$$m(x) = \begin{cases} m_l(x) & , x \leq \tau \\ m_r(x) & , x > \tau \end{cases} ,$$

for some $m_l, m_r : \mathbb{R} \to \mathbb{R}$, which are continuous in a punctured ε-neighborhood $U_\varepsilon(\tau) \setminus \{\tau\}$ of τ, with $m_l(\tau-) \neq m_r(\tau+)$ and

$$\left| m_l(\tau-) - \frac{\alpha+\beta}{2} \right| > 0 \quad \text{and} \quad \left| m_r(\tau+) - \frac{\alpha+\beta}{2} \right| > 0 \,. \tag{6.1}$$



(B2) For all sequences $(x_n)_{n\in\mathbb{N}}$ and $(x'_n)_{n\in\mathbb{N}}$ with $x_n \nearrow \tau$ and $x'_n \searrow \tau$, the sequences of regular conditional distributions $(\mathbb{P}_{Y|X}(x_n,\cdot))_{n\in\mathbb{N}}$ and $(\mathbb{P}_{Y|X}(x'_n,\cdot))_{n\in\mathbb{N}}$ are weakly convergent to measures denoted by $\mathbb{P}_{Y|X}(\tau-,\cdot)$ and $\mathbb{P}_{Y|X}(\tau+,.)$, respectively.

Remark 6.1 Recall that for a measurable function $g:\mathbb{R}\to\mathbb{R}$ the regular conditional distribution of $g\circ Y$ given $X=x$ is given by

$$\mathbb{P}_{g(Y)|X}(x,B)=\mathbb{P}_{Y|X}(x,g^{-1}(B)) \quad \text{for } x\in\mathbb{R},\ B\in\mathscr{B}(\mathbb{R})\,,$$

(cf. [GS77, Example 5.3.11]). If, in addition to the condition (B2), g is continuous, it follows by the CMT that

$$\begin{aligned}
\mathbb{P}_{g(Y)|X}(x_n,\cdot)&=\mathbb{P}_{Y|X}(x_n,g^{-1}(\cdot))\xrightarrow{w}\mathbb{P}_{Y|X}(\tau-,g^{-1}(\cdot))=:\mathbb{P}_{g(Y)|X}(\tau-,\cdot)\\
\mathbb{P}_{g(Y)|X}(x'_n,\cdot)&=\mathbb{P}_{Y|X}(x'_n,g^{-1}(\cdot))\xrightarrow{w}\mathbb{P}_{Y|X}(\tau+,g^{-1}(\cdot))=:\mathbb{P}_{g(Y)|X}(\tau+,\cdot)\,,
\end{aligned}$$

for all sequences $x_n \nearrow \tau$ and $x'_n \searrow \tau$.

Remark 6.2 Suppose that X and Y have a joint density function $f_{(X,Y)}$ with respect to Lebesgue-measure λ^2 on $\mathbb{R}^2$. Then by [Wit85, Proposition 1.126] the regular conditional distribution $\mathbb{P}_{Y|X}$ has a λ-density function $f_{Y|X}$ with

$$f_{Y|X}(x,y)=\frac{f_{(X,Y)}(x,y)}{\int_{\mathbb{R}} f_{(X,Y)}(x,y)\lambda(\mathrm{d}y)}$$

for $\mathbb{P}_X$-almost all $x\in\mathbb{R}$. If in addition there is an ε-neighborhood $U_\varepsilon(\tau)$ of τ such that for each $x\in U_\varepsilon(\tau)$ the marginal density function

$$f_X(x)=\int_{\mathbb{R}} f_{(X,Y)}(x,y)\lambda(\mathrm{d}y)$$

is strictly positive, then for each $x \in U_\varepsilon(\tau)$ the conditional probability $\mathbb{P}_{Y|X}(x,\cdot)$ can be explicitly specified by

$$\mathbb{P}_{Y|X}(x,A) = \int_A f_{Y|X}(x,y)\lambda(\,\mathrm{d}y) \quad \text{for all } A \in \mathscr{B}(\mathbb{R})\,.$$

Now assume there exists a measurable function $\tilde{f}: \mathbb{R}^2 \to \mathbb{R}_+$ such that $\int_{\mathbb{R}} \tilde{f}(x,y)\lambda(\,\mathrm{d}y) = 1$ for all $x \in U_\varepsilon(\tau)$. If for each $x_n \nearrow \tau$

$$\lim_{n\to\infty} f_{Y|X}(x_n,y) = \tilde{f}(\tau,y)$$

for λ-almost all $y \in \mathbb{R}$, then, as a consequence of Scheffés theorem ([Bil68, p. 224]), we have that

$$\lim_{n\to\infty} \mathbb{P}_{Y|X}(x_n,A) = \lim_{n\to\infty} \int_A f_{Y|X}(x_n,y)\lambda(\,\mathrm{d}y) = \int_A \tilde{f}(\tau,y)\lambda(\,\mathrm{d}y) =: \mathbb{P}_{Y|X}(\tau-,A)$$

for all $A \in \mathscr{B}(\mathbb{R})$. With the respective conditions for $x_n \searrow \tau$ the assumption of $(B2)$ is satisfied for the measures $\mathbb{P}_{Y|X}(\tau-,\cdot)$ and $\mathbb{P}_{Y|X}(\tau+,\cdot)$.

We will now collect some examples. The simplest form which the regression function m can assume under condition $(B1)$ is that it is a decision tree itself. This case has often been discussed, as for example in [Kos08, Section 14.5.1], [SS11a] or [SS11b, section 5.1], where the respective authors additionally assume that X and ϵ are independent.

Example 6.3 (Stump model, X and ϵ independent) If we assume that $m_l = \alpha$ and $m_r = \beta$, with $\alpha \neq \beta$, then

$$m(x) = \alpha \mathbb{1}_{(-\infty,\tau]}(x) + \beta \mathbb{1}_{(\tau,\infty)}(x)\,,$$

and we check at once that conditions (A1) and (B1) are fulfilled and $\|m\| < \infty$. If in addition, F is continuous and $\|V\| < \infty$, then Corollary 5.5 implies $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$, as $n \to \infty$. In the special case where X and ϵ are assumed to be independent, the regular conditional distribution and conditional variance of Y given $X = x$ can be specified as follows. For all $x \in (-\infty, \tau)$ and $B \in \mathscr{B}(\mathbb{R})$ by [GS77, Propositions 5.3.12

and 5.3.22] there would be

$$\begin{aligned}
\mathbb{P}_{Y|X}(x,B) &= \mathbb{E}\left(\mathbb{1}_{Y\in B}|X=x\right)\\
&= \mathbb{E}\left(\mathbb{1}_{(m(X)+\epsilon)\in B}|X=x\right)\\
&= \mathbb{E}\left(\mathbb{1}_{(m(x)+\epsilon)\in B}\right)\\
&= \mathbb{E}\left(\mathbb{1}_{(\alpha+\epsilon)\in B}\right)\\
&= \mathbb{P}\circ(\epsilon+\alpha)^{-1}(B)\,.
\end{aligned}$$

Similarly, for all $x\in(\tau,\infty)$ and $B\in\mathscr{B}(\mathbb{R})$, we have that $\mathbb{P}_{Y|X}(x,B)=\mathbb{P}\circ(\epsilon+\beta)^{-1}(B)$. Thus, condition (B2) is fulfilled, where $\mathbb{P}_{Y|X}(\tau+,B)=\mathbb{P}\circ(\epsilon+\beta)^{-1}(B)$ and $\mathbb{P}_{Y|X}(\tau-,B)=\mathbb{P}\circ(\epsilon+\alpha)^{-1}(B)$ for all $B\in\mathscr{B}(\mathbb{R})$. Furthermore, use [GS77, Propositions 5.3.22] to see that

$$V(x)=\mathbb{E}\left(\epsilon^2|X=x\right)=\mathbb{E}\left(Y-m(x)\right)^2=\mathbb{E}\left(Y-\alpha\right)^2\mathbb{1}_{(-\infty,\tau]}(x)+\mathbb{E}\left(Y-\beta\right)^2\mathbb{1}_{(\tau,\infty)}(x)$$

for $\mathbb{P}_X$-almost all $x\in\mathbb{R}$. Assumption (A3) then follows from $\mathbb{E}Y^2<\infty$. Moreover, by (2.10)

$$\mathbb{E}\left(\epsilon^2\right)=\int V(x)F(\mathrm{d}x)=\mathbb{E}\left(Y-\alpha\right)^2F(\tau)+\mathbb{E}\left(Y-\beta\right)^2\bar{F}(\tau)\,.$$

Example 6.4 If $\mathbb{P}_{Y|X}(x,\cdot)=\mathscr{N}\left(\mu(x),\sigma^2(x)\right)$ for functions $\mu:\mathbb{R}\to\mathbb{R}$ and $\sigma:\mathbb{R}\to(0,\infty)$, then

$$m(x)=\int y\mathbb{P}_{Y|X}(x,\mathrm{d}y)=\mathbb{E}\mathscr{N}\left(\mu(x),\sigma^2(x)\right)=\mu(x)$$

for $\mathbb{P}_X$-almost all $x\in\mathbb{R}$. Thus, m satisfies condition (B1) if and only if μ satisfies (B1). Analogously, $V(x)=\sigma^2(x)$ for $\mathbb{P}_X$-almost all $x\in\mathbb{R}$ and if, in addition to m, σ is continuous in $U_\varepsilon(\tau)\setminus\{\tau\}$ then condition (A3) is fulfilled. The conditional probability density function for $\mathbb{P}_{Y|X}(x,\cdot)$ is of the form

$$f_{Y|X}(x,y)=\frac{1}{\sqrt{2\pi\sigma^2(x)}}\exp\left\{-\frac{1}{2\sigma^2(x)}\left(y-\mu(x)\right)^2\right\}$$

for $\mathbb{P}_X$-almost all $x\in\mathbb{R}$. If we additionally assume that X is absolutely continuous with λ-density $f_X(x)>0$ for all $x\in U_\varepsilon(\tau)$ then, by the considerations from Remark 6.2, condition (B2) is satisfied with $\mathbb{P}_{Y|X}(\tau-,\mathrm{d}y)=\mathscr{N}\left(\mu(\tau-),\sigma^2(\tau-)\right)$ and $\mathbb{P}_{Y|X}(\tau+,\mathrm{d}y)=$

$\mathcal{N}\left(\mu(\tau+),\sigma^2(\tau+)\right)$. Moreover, if $\|\mu\|,\|\sigma\|<\infty$, then by Corollary 5.5, $(\tau_n,\alpha_n,\beta_n)\xrightarrow{\mathbb{P}}(\tau,\alpha,\beta)$.

6.2 Weak convergence of the rescaled process

Define the rescaled process

$$Z_n(t,a,b) := n\left\{S_n(\tau+n^{-1}t,\alpha+n^{-\frac{1}{2}}a,\beta+n^{-\frac{1}{2}}b)-S_n(\tau,\alpha,\beta)\right\}. \tag{6.2}$$

We inspect (2.13) to derive

$$Z_n(t,a,b) = 2(\beta-\alpha)Z_n^{(1)}(t)+Z_n^{(2)}(a)+Z_n^{(3)}(b)+R_n(t,a,b)\,,$$

where we define

$$\begin{aligned}
Z_n^{(1)}(t) &= \sum_{i=1}^n\left(\mathbb{1}_{X_i\le\tau+n^{-1}t}-\mathbb{1}_{X_i\le\tau}\right)\left(Y_i-\frac{\alpha+\beta}{2}\right)\\
Z_n^{(2)}(a) &= 2an^{-\frac{1}{2}}\sum_{i=1}^n\mathbb{1}_{\{X_i\le\tau\}}(\alpha-Y_i)+a^2F_n(\tau)\\
Z_n^{(3)}(b) &= 2bn^{-\frac{1}{2}}\sum_{i=1}^n\mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)+b^2\bar{F}_n(\tau)\\
R_n(t,a,b) &= \sum_{i=1}^n\left(\mathbb{1}_{X_i\le\tau+n^{-1}t}-\mathbb{1}_{X_i\le\tau}\right)\left(2n^{-\frac{1}{2}}\left[(a\alpha-b\beta)+(b-a)Y_i\right]+n^{-1}\left(a^2-b^2\right)\right).
\end{aligned} \tag{6.3}$$

As we can see it is possible to find a decomposition of Z_n into processes $2(\beta-\alpha)Z_n^{(1)}$, $Z_n^{(2)}$ and $Z_n^{(3)}$, each depends only on t, a and b, respectively, and a term $R_n(t,a,b)$ which depends on t, a and b. It will become apparent that the sum $2(\beta-\alpha)Z_n^{(1)}(t)+Z_n^{(2)}(a)+Z_n^{(3)}(b)$ determines the limit distribution of Z_n. On the other hand we will show that the remainder process $R_n(t,a,b)$ uniformly converges to zero in probability as n tends to ∞.

Definition 6.5 Let $(Z^{(1)}(t))_{t\in\mathbb{R}}\in D(\mathbb{R})$ be a two-sided compound Poisson process

$$Z^{(1)}(t) = \begin{cases}\sum\limits_{i=1}^{N_1(t)}\xi_i' & ,t\ge 0\\ \sum\limits_{i=1}^{N_2(-t)}\xi_i & ,t<0\end{cases},$$

i. e. a composition of two independent positive half-line compound Poisson processes. The continuous-time processes $(N_1(t))_{t\geq 0}$ and $(N_2(-t))_{t<0}$ are independent Poisson processes with intensities $F'_+(\tau)$ and $F'_-(\tau)$, respectively. The random variables $(\xi'_i)_{i\in\mathbb{N}}$ and $(\xi_i)_{i\in\mathbb{N}}$, independent of N_1 and N_2, are i. i. d. copies of $\mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\cdot)$ and $\mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau-,\cdot)$, named jump size distributions (cf. Remark 6.1). Additionally, let W and W' be two independent random variables with $W, W' \sim \mathcal{N}(0,1)$. We define the processes $(Z^{(2)}(a))_{a\in\mathbb{R}} \in D(\mathbb{R})$ and $(Z^{(3)}(b))_{b\in\mathbb{R}} \in D(\mathbb{R})$ by

$$
\begin{aligned}
Z^{(2)}(a) &:= 2\sqrt{T(\tau)}Wa + F(\tau)a^2 \\
Z^{(3)}(b) &:= 2\sqrt{\bar{T}(\tau)}W'b + \bar{F}(\tau)b^2 \,,
\end{aligned}
$$

where $T(\tau) = \mathbb{E}(\mathbb{1}_{X\leq\tau}(Y-\alpha)^2)$ and $\bar{T}(\tau) = \mathbb{E}(\mathbb{1}_{X>\tau}(Y-\beta)^2)$. We introduce a stochastic process $Z \in D(\mathbb{R}^3)$ with

$$
Z(t,a,b) = 2(\beta-\alpha)Z^{(1)}(t) + Z^{(2)}(a) + Z^{(3)}(b) \,,
$$

where $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$ are independent.

We note that $Z^{(2)}$ and $Z^{(3)}$ are simple Gaussian processes with mean functions $m(a) = a^2F(\tau)$ and $m(b) = b^2\bar{F}(\tau)$, respectively, and covariance functions $k(a,a') = 4aa'T(\tau)$ and $k(b,b') = 4bb'\bar{T}(\tau)$, respectively. Now we specify the expectations of the jump size distributions of the process $Z^{(1)}$.

Lemma 6.6 Let assumptions (A1)-(A3) and (B1)-(B2) hold. If ξ' and ξ are real valued random variables with $\xi' \sim \mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\cdot)$ and $\xi \sim \mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau-,\cdot)$, then

$$
\mathbb{E}\left(\xi'\right) = m_r(\tau+) - \frac{\alpha+\beta}{2} \quad \text{and} \quad \mathbb{E}\left(\xi\right) = \frac{\alpha+\beta}{2} - m_l(\tau-) \,.
$$

Furthermore,

$$
2(\beta-\alpha)\mathbb{E}\left(\xi'\right) > 0 \quad \text{and} \quad 2(\beta-\alpha)\mathbb{E}\left(\xi\right) > 0 \,.
$$

Proof. First consider the random variable $\xi' \sim \mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\cdot)$. From assumptions (A3) and (B1) there exists an $\varepsilon > 0$ such that

$$
\sup_{x\in(\tau,\tau+\varepsilon)} V(x) = \sup_{x\in(\tau,\tau+\varepsilon)} \int (y-m(x))^2\,\mathbb{P}_{Y|X}(x,\mathrm{d}y) < \infty
$$

and $\sup_{x\in(\tau,\tau+\varepsilon)}|m(x)| < \infty$. Since $y^2 \le 2((y-m(x))^2 + m(x)^2)$, we also have

$$\sup_{x\in(\tau,\tau+\varepsilon)} \int y^2 \mathbb{P}_{Y|X}(x,\mathrm{d}y) < \infty\,.$$

By [GS77, Corollary 1.14.8] then the family of conditional distributions $\big(\mathbb{P}_{Y|X}(x,\cdot)\big)_{x\in(\tau,\tau+\varepsilon)}$ is uniformly integrable and by [WM95, Proposition 5.92] from condition (B2) it follows that for a sequence $x_m \searrow \tau$ then

$$\lim_{m\to\infty} \int y\mathbb{P}_{Y|X}(x_m,\mathrm{d}y) = \int y\mathbb{P}_{Y|X}(\tau+,\mathrm{d}y)\,. \tag{6.4}$$

Now define

$$E := \left\{x\in\mathbb{R};\ \int y\ \mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(x,\mathrm{d}y) = \mathbb{E}\left(Y - \frac{\alpha+\beta}{2}|X = x\right)\right\}\,.$$

From [GS77, Proposition 5.3.12] it follows that $\mathbb{P}_X(E) = 1$. By condition (A2) and A.6 we even find a sequence $(x_n)_{n\in\mathbb{N}} \subseteq E$ with $x_n \searrow \tau$. Application of Equation (6.4), condition (B1) and basic properties of conditional expectations from [GS77, Proposition 5.3.12] and [WM95, Proposition 1.120] yields

$$\begin{aligned}
\mathbb{E}(\xi') &= \int y\ \mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\mathrm{d}y)\\
&= \lim_{x_m\searrow\tau} \int y\ \mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(x_m,\mathrm{d}y)\\
&= \lim_{x_m\searrow\tau} \mathbb{E}\left(Y - \frac{\alpha+\beta}{2}|X = x_m\right)\\
&= \lim_{x_m\searrow\tau} \mathbb{E}(Y|X = x_m) - \frac{\alpha+\beta}{2}\\
&= m_r(\tau+) - \frac{\alpha+\beta}{2}\,.
\end{aligned}$$

From Lemma 2.1(v) we find that $2(\beta-\alpha)\mathbb{E}(\xi') \ge 0$. Since $\alpha \ne \beta$, by (A1), and $|\mathbb{E}(\xi')| > 0$, by (B1), we follow that $2(\beta-\alpha)\mathbb{E}(\xi') > 0$. Similar considerations applied to $\xi \sim \mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau+,\cdot)$ yield

$$2(\beta-\alpha)\mathbb{E}(\xi) = 2(\beta-\alpha)\left(\frac{\alpha+\beta}{2} - m_l(\tau-)\right) > 0\,. \qquad \square$$

The rest of this section covers the proof of the following theorem.

Theorem 6.7 If (A1)-(A3) and (B1)-(B2) hold, then

$$Z_n \xrightarrow{\mathscr{L}} Z \quad \text{in } D(\mathbb{R}^3)\,,$$

as $n \to \infty$.

At first we will consider each of the sequences of processes $Z_n^{(1)}$, $Z_n^{(2)}$ and $Z_n^{(3)}$ separately. But to follow that $Z_n \xrightarrow{\mathscr{L}} Z$ in the multivariate Skorokhod space $D(\mathbb{R}^3)$ it is not sufficient only to proof convergence in $D(\mathbb{R})$ for each of the processes. To solve this problem we will make a detour via convergence in the Skorokhod product space D_3 in Proposition 6.13. The process Z_n will then be identified as an image of a continuous map Φ from D_3 to $D(\mathbb{R}^3)$. An application of the CMT then provides Theorem 6.7.

Lemma 6.8 Let $(\Gamma(t))_{t\in\mathbb{R}}$ be a two sided compound Poisson process

$$\Gamma(t) = \begin{cases} \sum_{i=1}^{N_1(t)} \xi_i' & , t \geq 0 \\ \sum_{i=1}^{N_2(-t)} \xi_i & , t < 0 \end{cases} \tag{6.5}$$

with intensities $\lambda, \lambda' > 0$, and jump size distributions μ_ξ and $\mu_{\xi'}$. For each finite set of points $T \subseteq \mathbb{R}$, ordered according to their size and denoted by $t_l < t_{l-1} < \ldots < t_1 < 0 \leq t_1' < \ldots < t_m'$, where $l, m \in \mathbb{N}_0$, the characteristic function of the fidis $\pi_T(\Gamma)$ is then given by

$$\begin{aligned}\varphi_{\pi_T(\Gamma)}(z) = \prod_{\iota=1}^{l} \exp\Big\{(t_{\iota-1} - t_\iota)\lambda\Big(\varphi_{\mu_\xi}(z_1 + \ldots + z_{l-\iota+1}) - 1\Big)\Big\} \\ \times \prod_{\nu=1}^{m} \exp\Big\{(t_\nu' - t_{\nu-1}')\lambda'\Big(\varphi_{\mu_{\xi'}}(z_{l+\nu} + \ldots + z_{l+m}) - 1\Big)\Big\}\end{aligned}$$

for all $z \in \mathbb{R}^{l+m}$, where we set $t_0 = t_0' = 0$.

Proof. Let $A \in \mathbb{R}^{l\times l}$ be an anti-bidiagonal matrix with entries 1 along and entries -1 below the anti-diagonal and let $A' \in \mathbb{R}^{m\times m}$ be a bidiagonal matrix with entries 1 along and entries −1 below the main diagonal. If M is a block diagonal matrix of the form

$$M = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & A' \end{pmatrix}, \tag{6.6}$$

then for the vector of increments $\Delta = \Big((\Gamma(t_\iota) - \Gamma(t_{\iota-1}))_{1\le\iota\le l}, \big(\Gamma(t'_\nu) - \Gamma(t'_{\nu-1})\big)_{1\le\nu\le m} \Big)^\top$ we can write $\Delta = M\pi_T(\Gamma)$. Furthermore, M is invertible and for each $z \in \mathbb{R}^{l+m}$ we have

$$\big(M^{-1}\big)^\top z = \begin{pmatrix} \big(A^{-1}\big)^\top & \mathbf{0} \\ \mathbf{0} & \big(A'^{-1}\big)^\top \end{pmatrix} z = \big((z_1 + \ldots + z_{l-\iota+1})_{1\le\iota\le l}, (z_{l+\nu} + \ldots + z_{l+m})_{1\le\nu\le m} \big)^\top . \tag{6.7}$$

Note that for each $t \in \mathbb{R}$, $\Gamma(t)$ is a compound Poisson random variable and its characteristic function is given by

$$\varphi_{\Gamma(t)}(z) = \begin{cases} \exp\Big(t\lambda' \Big(\varphi_{\mu_{\xi'}}(z) - 1 \Big) \Big) & , t \ge 0 \\ \exp\Big(-t\lambda \Big(\varphi_{\mu_{\xi}}(z) - 1 \Big) \Big) & , t < 0 \end{cases}$$

for each $z \in \mathbb{R}$; see for instance [App04, Example 1.3.10 together with Proposition 1.2.11]. Furthermore, a Poisson process is a Lévy process [App04, Propositon 1.3.11], so that in particular its increments are independent and stationary and we can use (6.7) to derive

$$\begin{aligned}
&\varphi_{\pi_T(\Gamma)}(z) \\
&= \varphi_{M^{-1}\Delta}(z) \\
&= \varphi_\Delta((M^{-1})^\top z) \\
&= \Bigg(\prod_{\iota=1}^{l} \varphi_{(\Gamma(t_\iota)-\Gamma(t_{\iota-1}))}(z_1 + \ldots + z_{l-\iota+1}) \Bigg) \Bigg(\prod_{\nu=1}^{m} \varphi_{\Gamma(t'_\nu)-\Gamma(t'_{\nu-1})}(z_{l+\nu} + \ldots + z_{l+m}) \Bigg) \\
&= \Bigg(\prod_{\iota=1}^{l} \varphi_{\Gamma(t_\iota - t_{\iota-1})}(z_1 + \ldots + z_{l-\iota+1}) \Bigg) \Bigg(\prod_{\nu=1}^{m} \varphi_{\Gamma(t'_\nu - t'_{\nu-1})}(z_{l+\nu} + \ldots + z_{l+m}) \Bigg) \\
&= \prod_{\iota=1}^{l} \exp\Big\{ (t_{\iota-1} - t_\iota)\lambda \Big(\varphi_{\mu_\xi}(z_1 + \ldots + z_{l-\iota+1}) - 1 \Big) \Big\} \\
&\qquad\qquad \times \prod_{\nu=1}^{m} \exp\Big\{ (t'_\nu - t'_{\nu-1})\lambda' \Big(\varphi_{\mu_{\xi'}}(z_{l+\nu} + \ldots + z_{l+m}) - 1 \Big) \Big\} .
\end{aligned} \tag{6.8}$$

□

Remark 6.9 (i) To be precise, the Poisson process $(N_2(-t))_{t<0}$ in (6.5) is càglàd. Therefore, the right continuous modification $N_2(-(t-))_{t<0}$ should be especially chosen to make Γ be càdlàg. But for reasons of readability we consciously avoid this.

(ii) In the first two equalities of (6.8) only general properties for characteristic functions were used. Hence, the proof already shows that it is sufficient to look at the increments to identify a process.

Lemma 6.10 Let $(X_i, Y_i)_{1\le i\le n}$, for $n \in \mathbb{N}$, be i. i. d. copies of a random vector $(X, Y) \sim \mathbb{P}_X \otimes \mathbb{P}_{Y|X}$, where $F(x) := \mathbb{P}_X((-\infty, x])$ satisfies assumption (A2) for some $\tau \in T_F$, and $\mathbb{P}_{Y|X}$ satisfies assumption (B2). Suppose that $g : (\mathbb{R}, \mathscr{B}(\mathbb{R})) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is continuous and $g \circ Y$ is integrable. Furthermore, suppose that there exists $\gamma > 0$ such that $\mathbb{E}\left(|g(Y)|\,|X = x\right) < \infty$ for $\mathbb{P}_X$-almost all $x \in U_\gamma(\tau)$. If $(\Gamma_n)_{n\in\mathbb{N}}$ is a sequence of continuous-time processes $(\Gamma_n(t))_{t\in\mathbb{R}}$ with

$$\Gamma_n(t) = \sum_{i=1}^{n} \left(\mathbb{1}_{X_i \le \tau + n^{-1}t} - \mathbb{1}_{X_i\le\tau}\right) g(Y_i)\,,$$

then

$$\Gamma_n \xrightarrow{\mathscr{L}} \Gamma \quad \text{in } D(\mathbb{R})\,,$$

as $n \to \infty$, where $(\Gamma(t))_{t\in\mathbb{R}}$ is a two sided compound Poisson process with intensities $F'_+(\tau)$ for $t \ge 0$ and $F'_-(\tau)$ for $t < 0$, and jump size distributions $\mathbb{P}_{g(Y)|X}(\tau+, \cdot)$ and $\mathbb{P}_{-g(Y)|X}(\tau-, \cdot)$.

Proof. Set $a > 0$. According to Proposition 3.5 we have to show that $\Gamma_n|_{[-a,a]} \xrightarrow{\mathscr{L}} \Gamma|_{[-a,a]}$. A necessary and sufficient condition for weak convergence in $D([-a, a])$ is the convergence of the finite-dimensional distributions of Γ_n together with tightness, see [JS03, Chapter VI, §3b, 3.20]. First apply Lévy's continuity theorem, cf. [Kal97, Theorem 4.3], and use characteristic functions to show that the fidis of Γ_n converge to those of Γ. Let $T \subseteq [-a, a]$ be a finite set. We use the terminology introduced in Lemma 6.8, i.e. $t_l < t_{l-1} < \ldots < t_1 < 0 \le t'_1 < \ldots < t'_m$. For the sake of brevity write $U_\iota := \left(\tau + n^{-1}t_\iota, \tau + n^{-1}t_{\iota-1}\right]$, $1 \le \iota \le l$, and $U'_\nu := \left(\tau + n^{-1}t'_{\nu-1}, \tau + n^{-1}t'_\nu\right]$, $1 \le \nu \le m$. Let $\mathbb{1}_U(X) \in \mathbb{R}^{l+m}$ be a random vector with

$$\mathbb{1}_U(X) = \left(-\mathbb{1}_{X\in U_1}, \ldots, -\mathbb{1}_{X\in U_l}, \mathbb{1}_{X\in U'_1}, \ldots, \mathbb{1}_{X\in U'_m}\right)\,,$$

where $\mathbb{1}_U(X_j) \in \mathbb{R}^{l+m}$ denotes the corresponding function for the realization X_j, $j \in \mathbb{N}$, of the random variable X. With the same definition for the matrix M from (6.6) the vector of increments $\Delta_n := M\pi_T(\Gamma_n)$ reads

$$\Delta_n = \Big((\Gamma_n(t_\iota) - \Gamma_n(t_{\iota-1}))_{1\le\iota\le l}, \big(\Gamma_n(t'_\nu) - \Gamma_n(t'_{\nu-1})\big)_{1\le\nu\le m}\Big)^\top = \sum_{j=1}^n g(Y_j)\mathbb{1}_U(X_j)\,.$$

For brevity, set $u := \left(M^{-1}\right)^\top z$. We compute the characteristic function of the fidis $\pi_T(\Gamma_n)$. For each $z \in \mathbb{R}^{l+m}$

$$\begin{aligned}
&\varphi_{\pi_T(\Gamma_n)}(z) \\
&= \varphi_{\Delta_n}(u) \\
&= \mathbb{E}\left(\exp\left(\mathrm{i}\,\langle u, \Delta_n\rangle\right)\right) \\
&= \mathbb{E}\left(\exp\left(\mathrm{i}\left\langle u, \sum_{j=1}^n g(Y_j)\mathbb{1}_U(X_j)\right\rangle\right)\right) \\
&= \mathbb{E}\left(\exp\left(\mathrm{i}\sum_{j=1}^n g(Y_j)\left\langle u, \mathbb{1}_U(X_j)\right\rangle\right)\right) \\
&= \mathbb{E}\left(\prod_{j=1}^n \exp\left(\mathrm{i}\, g(Y_j)\left\langle u, \mathbb{1}_U(X_j)\right\rangle\right)\right) \\
&\overset{(X_j,Y_j)\ i.i.d.}{=} \left[\mathbb{E}\left(\exp\left(\mathrm{i}\, g(Y)\,\langle u, \mathbb{1}_U(X)\rangle\right)\right)\right]^n \\
&= \left[\mathbb{E}\left(\sum_{k=0}^\infty \frac{\mathrm{i}^k g(Y)^k \langle u, \mathbb{1}_U(X)\rangle^k}{k!}\right)\right]^n \\
&= \left[1 + \mathbb{E}\left(\sum_{k=1}^\infty \frac{\mathrm{i}^k g(Y)^k \langle u, \mathbb{1}_U(X)\rangle^k}{k!}\right)\right]^n\,.
\end{aligned} \tag{6.9}$$

Note that for complex sequences $(c_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} c_n = c$ one gets $\lim_{n\to\infty}\left(1 + \frac{c_n}{n}\right)^n = \exp(c)$. According to Lemma 6.8 we need to show the pointwise convergence of

$$c_n := n\mathbb{E}\left(\sum_{k=1}^\infty \frac{\mathrm{i}^k g(Y)^k \langle u, \mathbb{1}_U(X)\rangle^k}{k!}\right)$$

towards

$$c := \sum_{\iota=1}^{l} (t_{\iota-1} - t_\iota) F'_-(\tau) \left(\varphi^{\tau-}_{-g(Y)} (z_1 + \ldots + z_{l-\iota+1}) - 1 \right) \\ + \sum_{\nu=1}^{m} \left(t'_\nu - t'_{\nu-1} \right) F'_+(\tau) \left(\varphi^{\tau+}_{g(Y)} (z_{l+\nu} + \ldots + z_{l+m}) - 1 \right),$$

as $n \to \infty$, where $\varphi^{\tau+}_{g(Y)}$ and $\varphi^{\tau-}_{-g(Y)}$ denote the characteristic functions of $\mathbb{P}_{g(Y)|X}(\tau+, \cdot)$ and $\mathbb{P}_{-g(Y)|X}(\tau-, \cdot)$, respectively. All indicator functions in c_n are indicator functions of disjoint sets. For a set A we have $\mathbb{1}_A^k = \mathbb{1}_A$, $(-\mathbb{1}_A)^k = (-1)^k \mathbb{1}_A$ and $\mathbb{1}_A^0 = 1$, $k \in \mathbb{N}$. Furthermore, $\mathbb{1}_{A \cap B} = 0$ if $A \cap B = \emptyset$. Hence, by expanding $\langle u, \mathbb{1}_U(X) \rangle^k$ with the multinomial theorem, all combinations of mixed powers of the terms vanish and

$$\langle u, \mathbb{1}_U(X) \rangle^k = -1^k \sum_{\iota=1}^{l} u_\iota^k \mathbb{1}_{X \in U_\iota} + \sum_{\nu=1}^{m} u_{l+\nu}^k \mathbb{1}_{X \in U'_\nu}.$$

We continue

$$\begin{aligned} c_n & \\ &= n\mathbb{E}\left(\sum_{k=1}^{\infty} \frac{\mathrm{i}^k g(Y)^k \left(-1^k \sum_{\iota=1}^{l} u_\iota^k \mathbb{1}_{X\in U_\iota} + \sum_{\nu=1}^{m} u_{l+\nu}^k \mathbb{1}_{X\in U'_\nu} \right)}{k!} \right) \\ &= n \left(\sum_{\iota=1}^{l} \mathbb{E}\left(\mathbb{1}_{X\in U_\iota} \sum_{k=1}^{\infty} \frac{\mathrm{i}^k (-g(Y))^k u_\iota^k}{k!} \right) + \sum_{\nu=1}^{m} \mathbb{E}\left(\mathbb{1}_{X\in U'_\nu} \sum_{k=1}^{\infty} \frac{\mathrm{i}^k g(Y)^k u_{l+\nu}^k}{k!} \right) \right) \\ &= \sum_{\iota=1}^{l} n \left[\mathbb{E}\left(\mathbb{1}_{X\in U_\iota} \exp(\mathrm{i}\,(-g(Y))\, u_\iota) \right) - \mathbb{E}\left(\mathbb{1}_{X\in U_\iota} \right) \right] \\ &\qquad + \sum_{\nu=1}^{m} n \left[\mathbb{E}\left(\mathbb{1}_{X\in U'_\nu} \exp(\mathrm{i}\, g(Y) u_{l+\nu}) \right) - \mathbb{E}\left(\mathbb{1}_{X\in U'_\nu} \right) \right]. \end{aligned}$$

We consider in more detail the expectation

$$\begin{aligned} &\mathbb{E}\left(\mathbb{1}_{X\in U_\iota} \exp(\mathrm{i}\,(-g(Y))\, u_\iota) \right) \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{x\in U_\iota} \exp(\mathrm{i}\,(-g(Y))\, u_\iota) (\mathbb{P}_X \otimes \mathbb{P}_{Y|X}) \, \mathrm{d}(x, y) \end{aligned}$$

$$= \int\limits_{U_\iota} \int\limits_{\mathbb{R}} \exp\left(\mathrm{i}\,(-g(y))\,u_\iota\right) \mathbb{P}_{Y|X}(x, \mathrm{d}y) \mathbb{P}_X(\mathrm{d}x)\,,$$

where we used [GS77, Proposition 1.8.10]. Because $\mathbb{P}_{-g(Y)|X}(x,B) = \mathbb{P}_{Y|X}(x, g^{-1}(-B))$ for all $x \in \mathbb{R}$ and $B \in \mathscr{B}(\mathbb{R})$ (cf. Remark 6.1), by the change of variable formula (cf. [GS77, Proposition 1.10.4]) then for each $x \in \mathbb{R}$

$$\int\limits_{\mathbb{R}} \exp\left(\mathrm{i}(-g(y))u_\iota\right) \mathbb{P}_{Y|X}(x, \mathrm{d}y) = \int\limits_{\mathbb{R}} \exp\left(\mathrm{i}zu_\iota\right) \mathbb{P}_{-g(Y)|X}(x, \mathrm{d}z) =: \varphi^x_{-g(Y)}(u_\iota)\,.$$

We observe that $\varphi^x_{-g(Y)}$ is the characteristic function of $\mathbb{P}_{-g(Y)|X}(x,\cdot)$. Furthermore, $\exp(\mathrm{i}zu_\iota)$ is continuous and $|\exp(izu_\iota)| \leq 1$, thus for each sequence $x_m \nearrow \tau$ by assumption (B2) and in consideration of Remark 6.1

$$\int\limits_{\mathbb{R}} \exp\left(\mathrm{i}(-g(y))u_\iota\right) \mathbb{P}_{Y|X}(x_m, \mathrm{d}y) \longrightarrow \int\limits_{\mathbb{R}} \exp\left(\mathrm{i}(-g(y))u_\iota\right) \mathbb{P}_{Y|X}(\tau-, \mathrm{d}y) =: \varphi^{\tau-}_{-g(Y)}(u_\iota)$$

and, therefore, $\varphi^{x_m}_{-g(Y)}(u) \to \varphi^{\tau-}_{-g(Y)}(u)$ for all $u \in \mathbb{R}$. Repeating the previous arguments leads to $\varphi^{x_m}_{g(Y)}(u) \to \varphi^{\tau+}_{g(Y)}(u)$ for all $u \in \mathbb{R}$ and $x_m \searrow \tau$. Using $\mathbb{1}_{X\in U_\iota} = \mathbb{1}_{X\in(\tau+n^{-1}t_\iota,\tau]} - \mathbb{1}_{X\in(\tau+n^{-1}t_{\iota-1},\tau]}$ and the corresponding result for $\mathbb{1}_{U'_\nu}$, then gives

$$\begin{aligned}
&c_n\\
&= \sum_{\iota=1}^{l} n\left[\int\limits_{U_\iota} \varphi^x_{-g(Y)}(u_\iota)F(\mathrm{d}x) - \big(F(\tau+n^{-1}t_{\iota-1}) - F(\tau+n^{-1}t_\iota)\big)\right]\\
&\qquad\qquad + \sum_{\nu=1}^{m} n\left[\int\limits_{U'_\nu} \varphi^x_{g(Y)}(u_\nu)F(\mathrm{d}x) - \big(F(\tau+n^{-1}t'_\nu) - F(\tau+n^{-1}t'_{\nu-1})\big)\right]\\
&= \sum_{\iota=1}^{l} n\left[\int\limits_{(\tau+n^{-1}t_\iota,\tau]} \varphi^x_{-g(Y)}(u_\iota)F(\mathrm{d}x) - \big(F(\tau) - F(\tau+n^{-1}t_\iota)\big)\right.\\
&\qquad\qquad \left. - \int\limits_{(\tau+n^{-1}t_{\iota-1},\tau]} \varphi^x_{-g(Y)}(u_\iota)F(\mathrm{d}x) - \big(F(\tau) - F(\tau+n^{-1}t_{\iota-1})\big)\right]
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{\nu=1}^{m} n\left[\int\limits_{(\tau,\tau+n^{-1}t'_\nu]} \varphi^x_{g(Y)}(u_\nu)F(\,\mathrm{d}x)-\big(F(\tau+n^{-1}t'_\nu)-F(\tau)\big)\right.\\
&\qquad\left.-\int\limits_{(\tau,\tau+n^{-1}t'_{\nu-1}]} \varphi^x_{g(Y)}(u_\nu)F(\,\mathrm{d}x)-\big(F(\tau+n^{-1}t'_{\nu-1})-F(\tau)\big)\right]\\
&=\sum_{\iota=1}^{l}\left[-t_\iota\left(\frac{\int\limits_{(\tau+n^{-1}t_\iota,\tau]}\varphi^x_{-g(Y)}(u_\iota)F(\,\mathrm{d}x)}{F(\tau)-F(\tau+n^{-1}t_\iota)}-1\right)\frac{F(\tau)-F(\tau+n^{-1}t_\iota)}{n^{-1}t_\iota}\right.\\
&\qquad\left.+t_{\iota-1}\left(\frac{\int\limits_{(\tau+n^{-1}t_{\iota-1},\tau]}\varphi^x_{-g(Y)}(u_\iota)F(\,\mathrm{d}x)}{F(\tau)-F(\tau+n^{-1}t_{\iota-1})}-1\right)\frac{F(\tau)-F(\tau+n^{-1}t_{\iota-1})}{n^{-1}t_{\iota-1}}\right]\\
&+\sum_{\nu=1}^{m}\left[t'_\nu\left(\frac{\int\limits_{(\tau,\tau+n^{-1}t'_\nu]}\varphi^x_{g(Y)}(u_\nu)F(\,\mathrm{d}x)}{F(\tau+n^{-1}t'_\nu)-F(\tau)}-1\right)\frac{F(\tau+n^{-1}t'_\nu)-F(\tau)}{n^{-1}t'_\nu}\right.\\
&\qquad\left.-t'_{\nu-1}\left(\frac{\int\limits_{(\tau,\tau+n^{-1}t'_{\nu-1}]}\varphi^x_{g(Y)}(u_\nu)F(\,\mathrm{d}x)}{F(\tau+n^{-1}t'_{\nu-1})-F(\tau)}-1\right)\frac{F(\tau+n^{-1}t'_{\nu-1})-F(\tau)}{n^{-1}t'_{\nu-1}}\right].
\end{aligned}
$$

Due to assumptions (A2), the results above and A.5 it follows that $\lim_{n\to\infty} c_n = c$. To proof tightness we use the moment criterion formulated in [FV09, Proposition 4.1]. Let $s < t < u$, then there exists an $n_0 = n_0(\varepsilon\wedge\gamma, a)\in\mathbb{N}$, such that for all $n \geq n_0$

$$
\begin{aligned}
&\mathbb{E}\left\{|\Gamma_n(t)-\Gamma_n(s)|\,|\Gamma_n(u)-\Gamma_n(t)|\right\}\\
&=\mathbb{E}\left\{\left|\sum_{i=1}^{n} g(Y_i)\mathbb{1}_{\{\tau+n^{-1}s<X_i\leq\tau+n^{-1}t\}}\right|\left|\sum_{i=1}^{n} g(Y_i)\mathbb{1}_{\{\tau+n^{-1}t<X_i\leq\tau+n^{-1}u\}}\right|\right\}\\
&\leq\mathbb{E}\left\{\sum_{i=1}^{n}|g(Y_i)|\,\mathbb{1}_{\{\tau+n^{-1}s<X_i\leq\tau+n^{-1}t\}}\sum_{i=1}^{n}|g(Y_i)|\,\mathbb{1}_{\{\tau+n^{-1}t<X_i\leq\tau+n^{-1}u\}}\right\}\\
&=\mathbb{E}\left\{\sum_{\substack{i,j=1\\ i\neq j}}^{n}|g(Y_i)|\,\mathbb{1}_{\{\tau+n^{-1}s<X_i\leq\tau+n^{-1}t\}}\left|g(Y_j)\right|\mathbb{1}_{\{\tau+n^{-1}t<X_j\leq\tau+n^{-1}u\}}\right\}\\
&\leq n^2\mathbb{E}\left\{|g(Y)|\,\mathbb{1}_{\{\tau+n^{-1}s<X\leq\tau+n^{-1}t\}}\right\}\mathbb{E}\left\{|g(Y)|\,\mathbb{1}_{\{\tau+n^{-1}t<X\leq\tau+n^{-1}u\}}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= n^2 \int \mathbb{E}\big(|g(Y)|\mathbb{1}_{\{\tau+n^{-1}s<X\le\tau+n^{-1}t\}}|X\big)\,\mathrm{d}\mathbb{P} \int \mathbb{E}\big(|g(Y)|\mathbb{1}_{\{\tau+n^{-1}t<X\le\tau+n^{-1}u\}}|X\big)\,\mathrm{d}\mathbb{P} \\
&= n^2 \int\limits_{(\tau+n^{-1}s,\tau+n^{-1}t]} \mathbb{E}(|g(Y)|\,|X=x)\,F(\mathrm{d}x) \int\limits_{(\tau+n^{-1}t,\tau+n^{-1}u]} \mathbb{E}(|g(Y)|\,|X=x)\,F(\mathrm{d}x) \\
&\le \|\mathbb{E}(|g(Y)|\,|X=\cdot)\|^2_{U_\gamma(\tau)}\,\overline{L}^2(t-s)(u-t) \\
&\le \|\mathbb{E}(|g(Y)|\,|X=\cdot)\|^2_{U_\gamma(\tau)}\,\overline{L}^2(u-s)^2\,. \qquad \square
\end{aligned}
$$

Corollary 6.11 If assumptions (A1)-(A3) and (B1)-(B2) hold, then

$$
Z_n^{(1)} \xrightarrow{\mathscr{L}} Z^{(1)} \quad \text{in } D(\mathbb{R})\,,
$$

as $n\to\infty$, where $Z^{(1)}$ is the process introduced in Definition 6.5.

Proof. Choose $\varepsilon>0$ such that $U_\varepsilon(\tau)$ meets the conditions in (B1) and (A2)-(A3). Using $|x|\le 1+x^2$, $x\in\mathbb{R}$, properties for conditional expectations (cf. [GS77, Proposition 5.2.20]) and assumption (B1) yields $\mathbb{P}_X$-almost surely

$$
\begin{aligned}
&\mathbb{E}\left(\left|Y-\frac{\alpha+\beta}{2}\right|\,|X=x\right) \\
&\le 1+\mathbb{E}\big((Y-m(X))^2\,|X=x\big)+\mathbb{E}\left(\left|m(X)-\frac{\alpha+\beta}{2}\right|\,|X=x\right) \\
&\le 1+\|V\|_{U_\varepsilon(\tau)}+\|m\|_{U_\varepsilon(\tau)}+\left|\frac{\alpha+\beta}{2}\right| \\
&<\infty\,.
\end{aligned}
$$

Now Lemma 6.10 implies the desired result. $\square$

Lemma 6.12 If assumption (A1) holds, then

$$
Z_n^{(2)} \xrightarrow{\mathscr{L}} Z^{(2)} \text{ and } Z_n^{(3)} \xrightarrow{\mathscr{L}} Z^{(3)} \quad \text{in } D(\mathbb{R})\,,
$$

as $n\to\infty$, where $Z^{(2)}$ and $Z^{(3)}$ are the processes introduced in 6.5.

Proof. Given $b>0$, let $\emptyset\ne T\subseteq[-b,b]$ be a finite set with ordered elements denoted by $a_1<a_2<\ldots<a_l$, $l\in\mathbb{N}$. At first we will show that

$$
V_n(a):=n^{-\frac{1}{2}}\sum_{i=1}^n \mathbb{1}_{X_i\le\tau}2(\alpha-Y_i)a \xrightarrow{\mathscr{L}} 2\sqrt{T(\tau)}Wa=:V(a) \quad \text{in } D([-b,b])\,,
$$

where $W \sim \mathcal{N}(0,1)$. Observe that $\Sigma := (4T(\tau)a_i a_j)_{1\leq i,j\leq l}$ is the covariance matrix of the random vector $2\sqrt{T(\tau)}W\ (a_1,\dots,a_l)^\top$. As in the proof of Lemma 6.10 the procedure now is to show the convergence of $\pi_T(V_n)$ to $\pi_T(V)$, on the one hand, where $\pi_T(V) = \mathcal{N}_l(0,\Sigma)$. On the other hand we have to show tightness of V_n. We have

$$\pi_T(V_n) = n^{-\frac{1}{2}}\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau} 2(\alpha - Y_i)\ (a_1,\dots,a_l)^\top,$$

where, by Lemma 2.1(iii), $(\mathbb{1}_{X_i\leq\tau}2(\alpha-Y_i)\ (a_1,\dots,a_l)^\top)_{i\in\mathbb{N}}$ is a sequence of i. i. d. mean-zero random vectors with covariance matrix Σ. Applying the multivariate central limit theorem yields

$$n^{-\frac{1}{2}}\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau} 2(\alpha - Y_i)\ (a_1,\dots,a_l)^\top \xrightarrow{\mathcal{L}} \mathcal{N}_l(0,\Sigma)\ .$$

To proof tightness we use the criterion formulated in [Bil68, Theorem 15.7]. Fix $s < t < u$ and note that $\mathbb{E}(\mathbb{1}_{X_j\leq\tau}(\alpha-Y_j)\mathbb{1}_{X_k\leq\tau}(\alpha-Y_k)) = 0$ when $j \neq k$. Then

$$\begin{aligned}
&\mathbb{E}\left(|V_n(t)-V_n(s)|\,|V_n(u)-V_n(t)|\right)\\
&= \mathbb{E}\left(\left|2(t-s)n^{-\frac{1}{2}}\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau}(\alpha-Y_i)\right|\left|2(u-t)n^{-\frac{1}{2}}\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau}(\alpha-Y_i)\right|\right)\\
&= n^{-1}4(t-s)(u-t)\mathbb{E}\left(\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau}(\alpha-Y_i)\right)^2\\
&= 4(t-s)(u-t)T(\tau)\\
&\leq 4(u-s)^2T(\tau)
\end{aligned}$$

and it follows that $V_n \xrightarrow{\mathcal{L}} V$ in $D([-b,b])$. Furthermore, by the SLLN for all $a \in \mathbb{R}$, $D_n(a) := F_n(\tau)a^2 \to F(\tau)a^2 =: D(a)$ almost surely as $n \to \infty$, where

$$\sup_{-b\leq a\leq b} |D_n(a) - D(a)| = \sup_{|a|\leq b}\left|(F_n(\tau)-F(\tau))\,a^2\right| \leq |b|^2\,|F_n(\tau)-F(\tau)| \longrightarrow 0$$

almost surely as $n \to \infty$. Slutsky's theorem then implies $(V_n, D_n) \xrightarrow{\mathcal{L}} (V,D)$ and CMT implies $Z_n^{(2)} \xrightarrow{\mathcal{L}} V + D = Z^{(2)}$. Using the same arguments it also follows that $Z_n^{(3)} \xrightarrow{\mathcal{L}} Z^{(3)}$.□

Proposition 6.13 If (A1)-(A3) and (B1)-(B2) hold, then

$$\left(Z_n^{(1)}, Z_n^{(2)}, Z_n^{(3)}\right) \xrightarrow{\mathscr{L}} \left(Z^{(1)}, Z^{(2)}, Z^{(3)}\right) \quad \text{in } (D_3, \mathscr{D}_3) \ ,$$

as $n \to \infty$.

Proof. We will use the procedure described in Proposition 3.7. As in the proof of Lemma 6.10 first use Lévy's continuity theorem to show the convergence of the finite dimensional distributions. Let $\emptyset \neq T \subseteq \mathbb{R}$ be a finite set of points, ordered according to their size and denoted by $t_l < t_{l-1} < \ldots < t_1 < 0 \leq t'_1 < \ldots < t'_m$, where $l, m \in \mathbb{N}_0$ and $t_0 = t'_0 = 0$. Set

$$\begin{aligned} t &= \left(t_l, \ldots, t_1, t'_1, \ldots, t'_m\right)^\top , \\ t^2 &= \left(t_l^2, \ldots, t_1^2, t_1'^2, \ldots, t_m'^2\right)^\top , \end{aligned}$$

and let $z = (u, v, w)^\top \in \mathbb{R}^{3(l+m)}$ be a vector composed of the vectors $u, v, w \in \mathbb{R}^{l+m}$. In particular, the vector of fidis of the process $Z^{(2)}$ then reads $\pi_T(Z^{(2)}) = 2\sqrt{T(\tau)} W t + F(\tau) t^2$, where $W \sim \mathscr{N}(0,1)$. Since W has characteristic function $\varphi_W(z) = \exp(-{}^1\!/\!{}_2\, z^2)$, $z \in \mathbb{R}$, then

$$\begin{aligned} \varphi_{\pi_T(Z^{(2)})}(v) &= \mathbb{E}\left(\exp\left(\mathrm{i} \left\langle v, \pi_T(Z^{(2)})\right\rangle\right)\right) \\ &= \exp\left(\mathrm{i} F(\tau) \left\langle v, t^2\right\rangle\right) \varphi_W\left(2\sqrt{T(\tau)}\, \langle v, t\rangle\right) \\ &= \exp\left(\mathrm{i} F(\tau) \left\langle v, t^2\right\rangle - 2T(\tau) \left\langle v, t\, t^\top v\right\rangle\right) \ . \end{aligned}$$

The previous arguments applied to the process $Z^{(3)}$ lead to the characteristic function

$$\varphi_{\pi_T(Z^{(3)})}(w) = \exp\left(\mathrm{i} \bar{F}(\tau) \left\langle w, t^2\right\rangle - 2\bar{T}(\tau) \left\langle w, t\, t^\top w\right\rangle\right) \ .$$

Because $Z^{(1)}$, $Z^{(2)}$ and $Z^{(3)}$ are independent we may determine the characteristic function of the vector of fidis $(\pi_T(Z^{(1)}), \pi_T(Z^{(2)}), \pi_T(Z^{(3)}))$:

$$
\begin{aligned}
&\varphi_{\left(\pi_T\left(Z^{(1)}\right),\pi_T\left(Z^{(2)}\right),\pi_T\left(Z^{(3)}\right)\right)}(z) \\
&= \varphi_{\pi_T(Z^{(1)})}(u)\,\varphi_{\pi_T(Z^{(2)})}(v)\,\varphi_{\pi_T(Z^{(3)})}(w) \\
&= \prod_{\iota=1}^{l} \exp\left\{(t_{\iota-1}-t_\iota)F'_-(\tau)\left(\varphi_{\mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau-,\cdot)}(u_1+\ldots+u_{l-\iota+1})-1\right)\right\} \\
&\qquad\qquad \times \prod_{\nu=1}^{m} \exp\left\{(t'_\nu-t'_{\nu-1})F'_+(\tau)\left(\varphi_{\mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\cdot)}(u_{l+\nu}+\ldots+u_{l+m})-1\right)\right\} \\
&\qquad \times \exp\left\{\mathrm{i}\,F(\tau)\left\langle v, t^2\right\rangle - 2T(\tau)\left\langle v, t\,t^\top v\right\rangle\right\}\exp\left\{\mathrm{i}\,\bar{F}(\tau)\left\langle w, t^2\right\rangle - 2\bar{T}(\tau)\left\langle w, t\,t^\top w\right\rangle\right\}\,,
\end{aligned}
\tag{6.10}
$$

where we used the results of Lemma 6.8. Now recall the definition of U_ι, U'_ν, $\mathbb{1}_U(X)$ and M from the proof of Lemma 6.10. Set $g(Y) = Y - \frac{\alpha+\beta}{2}$ and the random vectors $q(Y), q'(Y) \in \mathbb{R}^{m+l}$, with $q(Y) = 2n^{-1/2}(\alpha - Y)t + n^{-1}t^2$ and $q'(Y) = 2n^{-1/2}(\beta - Y)t + n^{-1}t^2$. Again $q(Y_j)$ and $q'(Y_j)$ denote the corresponding function for the random variables Y_j, $j \in \mathbb{N}$, of Y. Let K be the block diagonal matrix of the form $K = \operatorname{diag}(M, E, E)$, where E denotes the identity matrix of size $l+m$. By the special form of K and M, K is invertible and $(K^{-1})^\top = \operatorname{diag}\left((M^{-1})^\top, E, E\right)$, where $(M^{-1})^\top$ was already computed in (6.7). To study the characteristic function of the vector $(\pi_T(Z_n^{(1)}), \pi_T(Z_n^{(2)}), \pi_T(Z_n^{(3)}))$, $n \in \mathbb{N}$, adapt the computation made in (6.9). Then for each $z = (u, v, w) \in \mathbb{R}^{3(l+m)}$

$$
\begin{aligned}
&\varphi_{\left(\pi_T\left(Z_n^{(1)}\right),\pi_T\left(Z_n^{(2)}\right),\pi_T\left(Z_n^{(3)}\right)\right)}(z) \\
&= \varphi_{\left(M\pi_T\left(Z_n^{(1)}\right),\pi_T\left(Z_n^{(2)}\right),\pi_T\left(Z_n^{(3)}\right)\right)}\left((K^{-1})^\top z\right) \\
&= \mathbb{E}\left(\exp\left(\mathrm{i}\left(\left\langle \sum_{j=1}^{n} g(Y_j)\mathbb{1}_U(X_j),\, (M^{-1})^\top u\right\rangle\right.\right.\right. \\
&\qquad\qquad\qquad + \left\langle \sum_{j=1}^{n} 2n^{-\frac{1}{2}}\mathbb{1}_{X_j\le\tau}(\alpha - Y_j)t + n^{-1}\mathbb{1}_{X_j\le\tau}t^2, v\right\rangle \\
&\qquad\qquad\qquad \left.\left.\left.+ \left\langle \sum_{j=1}^{n} 2n^{-\frac{1}{2}}\mathbb{1}_{X_j>\tau}(\beta - Y_j)t + n^{-1}\mathbb{1}_{X_j>\tau}t^2, w\right\rangle\right)\right)\right) \\
&= \mathbb{E}\left(\exp\left(\mathrm{i}\sum_{j=1}^{n}\left(g(Y_j)\left\langle \mathbb{1}_U(X_j), (M^{-1})^\top u\right\rangle + \mathbb{1}_{X_j\le\tau}\left\langle q(Y_j), v\right\rangle + \mathbb{1}_{X_j>\tau}\left\langle q'(Y_j), w\right\rangle\right)\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \left[\mathbb{E}\left(\exp\left(\mathrm{i}\left(g(Y)\left\langle \mathbb{1}_U(X), (M^{-1})^\top u\right\rangle + \mathbb{1}_{X\le\tau}\left\langle q(Y),v\right\rangle + \mathbb{1}_{X>\tau}\left\langle q'(Y),w\right\rangle\right)\right)\right)\right]^n \\
&= \left[1+\mathbb{E}\left(\sum_{k=1}^{\infty}\frac{\mathrm{i}^k}{k!}\left\{g(Y)\left\langle \mathbb{1}_U(X), (M^{-1})^\top u\right\rangle + \mathbb{1}_{X\le\tau}\left\langle q(Y),v\right\rangle + \mathbb{1}_{X>\tau}\left\langle q'(Y),w\right\rangle\right\}^k\right)\right]^n .
\end{aligned}
$$

Splitting $\mathbb{1}_{X\le\tau} = \mathbb{1}_{X\le\tau+n^{-1}t_l} + \sum_{j=1}^{l}\mathbb{1}_{X\in U_j}$ and $\mathbb{1}_{X>\tau} = \sum_{j=1}^{m}\mathbb{1}_{X\in U_j'} + \mathbb{1}_{X>\tau+n^{-1}t_m'}$ leads to sums of indicator functions of mutually disjoint sets. By expanding the term in braces with the multinomial theorem we observe

$$
\begin{aligned}
&\left\{g(Y)\left\langle \mathbb{1}_U(X), (M^{-1})^\top u\right\rangle + \mathbb{1}_{X\le\tau}\left\langle q(Y),v\right\rangle + \mathbb{1}_{X>\tau}\left\langle q'(Y),w\right\rangle\right\}^k \\
&= \mathbb{1}_{X\le\tau+n^{-1}t_l}\left\langle q(Y),v\right\rangle^k + \sum_{j=1}^{l}\mathbb{1}_{X\in U_j}\left(-g(Y)(u_1+\ldots u_{l-j+1}) + \left\langle q(Y),v\right\rangle\right)^k \\
&\qquad + \sum_{j=1}^{m}\mathbb{1}_{X\in U_j'}\left(g(Y)(u_{l+j}+\ldots u_{l+m}) + \left\langle q'(Y),w\right\rangle\right)^k + \mathbb{1}_{X>\tau+n^{-1}t_m'}\left\langle q'(Y),w\right\rangle^k .
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&n\left(\left(\varphi_{\left(\pi_T\left(Z_n^{(1)}\right),\pi_T\left(Z_n^{(2)}\right),\pi_T\left(Z_n^{(3)}\right)\right)}(z)\right)^{n^{-1}} - 1\right) \\
&= n\mathbb{E}\left(\exp\left(\mathrm{i}\ \mathbb{1}_{X\le\tau+n^{-1}t_l}\left\langle q(Y),v\right\rangle\right)-1\right) \\
&\qquad + \sum_{j=1}^{l} n\mathbb{E}\left(\mathbb{1}_{X\in U_j}\left(\exp\left(\mathrm{i}\left[-g(Y)\left(u_1+\ldots+u_{l-j+1}\right)+\left\langle q(Y),v\right\rangle\right]\right)-1\right)\right) \\
&\qquad + \sum_{j=1}^{m} n\mathbb{E}\left(\mathbb{1}_{X\in U_j'}\left(\exp\left(\mathrm{i}\left[g(Y)\left(u_{l+j}+\ldots+u_{l+m}\right)+\left\langle q'(Y),w\right\rangle\right]\right)-1\right)\right) \\
&\qquad\qquad + n\mathbb{E}\left(\exp\left(\mathrm{i}\ \mathbb{1}_{X>\tau+n^{-1}t_m'}\left\langle q'(Y),w\right\rangle\right)-1\right) . \qquad (6.11)
\end{aligned}
$$

For the first summand on the right-hand side of Equation (6.11) we use the Taylorexpansion of characteristic functions (see [Bre92, Proposition 11.7]) to show

$$
\begin{aligned}
& n\mathbb{E}\left(\exp\left(\mathrm{i}\,\left\langle \mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y), v\right\rangle\right)-1\right)\\
&= n\mathbb{E}\left(\exp\left(\mathrm{i}\,\left\langle n^{\frac{1}{2}}\mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y), n^{-\frac{1}{2}}v\right\rangle\right)-1\right)\\
&= n\left(\mathrm{i}\,\mathbb{E}\left\langle \mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y), v\right\rangle - \frac{1}{2}\mathbb{E}\left\langle \mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y), v\right\rangle^2 + \delta(n^{-\frac{1}{2}}v)\left\|n^{-\frac{1}{2}}v\right\|^2\right)\\
&= n\left(\mathrm{i}\,\left\langle \mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y)\right), v\right\rangle - \frac{1}{2}\left\langle v, \mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y)q(Y)^{\top}\right)v\right\rangle\right) + \delta(n^{-\frac{1}{2}}v)\left\|v\right\|^2 ,
\end{aligned}
$$

where δ denotes a function, such that $\delta(n^{-\frac{1}{2}}v) \to 0$, as $n \to \infty$. Note that

$$
\begin{aligned}
\mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y)\right) &= 2n^{-\frac{1}{2}}\mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l}(\alpha - Y)\right)t + n^{-1}\mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l}\right)t^2\\
&= 2n^{-\frac{1}{2}}\mathbb{E}\left(\mathbb{1}_{\tau+n^{-1}t_l<X\le\tau}(Y-\alpha)\right)t + n^{-1}\mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l}\right)t^2\\
&= 2n^{-\frac{1}{2}}\int\limits_{(\tau+n^{-1}t_l,\tau]}(m(x)-\alpha)F(\mathrm{d}x)\,t + n^{-1}F\left(\tau+n^{-1}t_l\right)t^2 ,
\end{aligned}
$$

where we used Lemma 2.1(iii) at the second step and Equation (2.3) at the last step. Furthermore, we compute the matrix

$$
\begin{aligned}
& \mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l} q(Y)q(Y)^{\top}\right)\\
&= 4n^{-1}\mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l}(\alpha-Y)^2\right)t\,t^{\top} + 2n^{-\frac{3}{2}}\mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l}(\alpha-Y)\right)\left(t\,(t^2)^{\top}+t^2\,t^{\top}\right)\\
&\qquad + 2n^{-2}\mathbb{E}\left(\mathbb{1}_{X\le\tau+n^{-1}t_l}\right)t^2\,(t^2)^{\top}\\
&= 4n^{-1}\left(\int\limits_{(-\infty,\tau+n^{-1}t_l]}V(x)F(\mathrm{d}x) + \int\limits_{(-\infty,\tau+n^{-1}t_l]}(m(x)-\alpha)^2F(\mathrm{d}x)\right)t\,t^{\top}\\
&\quad + 2n^{-\frac{3}{2}}\int\limits_{(\tau+n^{-1}t_l,\tau]}(m(x)-\alpha)F(\mathrm{d}x)\,\left(t\,(t^2)^{\top}+t^2\,t^{\top}\right) + 2n^{-2}F\left(\tau+n^{-1}t_l\right)\,t^2\,(t^2)^{\top},
\end{aligned}
$$

where we used the decomposition from (2.10). By using assumptions (A2)-(A3) and (B1) and applying A.5 and the monotone convergence theorem, we derive convergence for the first term in Equation (6.11)

$$
\begin{aligned}
&n\mathbb{E}\left(\exp\left(\mathrm{i}\,\mathbb{1}_{X\le\tau+n^{-1}t_l}\,\langle q(Y),v\rangle\right)-1\right)\\
&=\mathrm{i}\left(2n^{\frac{1}{2}}\int\limits_{(\tau+n^{-1}t_l,\tau]}(m(x)-\alpha)F(\,\mathrm{d}x)\,\langle t,v\rangle+F\left(\tau+n^{-1}t_l\right)\left\langle t^2,v\right\rangle\right)\\
&\qquad-2\left(\int\limits_{(-\infty,\tau+n^{-1}t_l]}V(x)F(\,\mathrm{d}x)+\int\limits_{(-\infty,\tau+n^{-1}t_l]}(m(x)-\alpha)^2F(\,\mathrm{d}x)\right)\left\langle v,t\,t^{\top}v\right\rangle\\
&\qquad-n^{-\frac{1}{2}}\int\limits_{(\tau+n^{-1}t_l,\tau]}(m(x)-\alpha)F(\,\mathrm{d}x)\left\langle v,\left(t\,(t^2)^{\top}+t^2\,t^{\top}\right)v\right\rangle\\
&\qquad-n^{-1}F\left(\tau+n^{-1}t_l\right)\left\langle v,t^2\,(t^2)^{\top}v\right\rangle+\delta(n^{-\frac{1}{2}}v)\,\|v\|^2\\
&\longrightarrow\mathrm{i}\,F(\tau)\left\langle t^2,v\right\rangle-2T(\tau)\left\langle v,t\,t^{\top}v\right\rangle\ ,
\end{aligned}
$$

as $n\to\infty$. With the same arguments as above we can handle the fourth term in Equation (6.11) and get

$$
n\mathbb{E}\left(\exp\left(\mathrm{i}\,\mathbb{1}_{X>\tau+n^{-1}t'_m}\left\langle q'(Y),w\right\rangle\right)-1\right)\longrightarrow\mathrm{i}\,\bar{F}(\tau)\left\langle t^2,w\right\rangle-2\bar{T}(\tau)\left\langle w,t\,t^{\top}w\right\rangle\ ,
$$

as $n\to\infty$. Consider the second summand on the right-hand side of Equation (6.11) and compute

$$
\begin{aligned}
&\sum_{j=1}^{l}n\mathbb{E}\left(\mathbb{1}_{X\in U_j}\left(\exp\left(\mathrm{i}\left[-g(Y)\left(u_1+\ldots+u_{l-j+1}\right)+\langle q(Y),v\rangle\right]\right)-1\right)\right)\\
&=\sum_{j=1}^{l}n\left\{\mathbb{E}\left(\mathbb{1}_{\tau+n^{-1}t_j<X\le\tau}\left(\exp\left(\mathrm{i}\left[-g(Y)\left(u_1+\ldots+u_{l-j+1}\right)+\langle q(Y),v\rangle\right]\right)-1\right)\right)\right.\\
&\qquad\left.-\mathbb{E}\left(\mathbb{1}_{\tau+n^{-1}t_{j-1}<X\le\tau}\left(\exp\left(\mathrm{i}\left[-g(Y)\left(u_1+\ldots+u_{l-j+1}\right)+\langle q(Y),v\rangle\right]\right)-1\right)\right)\right\}
\end{aligned}
$$

$$= \sum_{j=1}^{l} n \Bigg\{ \int\limits_{\left(\tau+n^{-1}t_j,\tau\right]} \int \exp\left(\mathrm{i}\left[-g(y)\left(u_1+\ldots+u_{l-j+1}\right)+\langle q(y),v\rangle\right]\right)\mathbb{P}_{Y|X}(x,\mathrm{d}y)F(\mathrm{d}x) - \left(F(\tau)-F(\tau+n^{-1}t_j)\right)$$
$$- \int\limits_{\left(\tau+n^{-1}t_{j-1},\tau\right]} \int \exp\left(\mathrm{i}\left[-g(y)\left(u_1+\ldots+u_{l-j+1}\right)+\langle q(y),v\rangle\right]\right)\mathbb{P}_{Y|X}(x,\mathrm{d}y)F(\mathrm{d}x)$$
$$+ \left(F(\tau)-F(\tau+n^{-1}t_{j-1})\right) \Bigg\}\,.$$

The inner integrals can be handled in the following way

$$\begin{aligned}
&\int \exp\left(\mathrm{i}\left[-g(y)\left(u_1+\ldots+u_{l-j+1}\right)+\langle q(y),v\rangle\right]\right)\mathbb{P}_{Y|X}(x,\mathrm{d}y)\\
&= \int \exp\left(\mathrm{i}\left[-g(y)\left(u_1+\ldots+u_{l-j+1}\right))\right]\right)\mathbb{P}_{Y|X}(x,\mathrm{d}y)\\
&\qquad + \int \exp\left(\mathrm{i}\left[-g(y)\left(u_1+\ldots+u_{l-j+1}\right)\right]\right)(\exp(\mathrm{i}\,\langle q(y),v\rangle)-1)\,\mathbb{P}_{Y|X}(x,\mathrm{d}y)\,,
\end{aligned}$$

whereas the second term

$$\begin{aligned}
&\left|\int \exp\left(\mathrm{i}\left[-g(y)\left(u_1+\ldots+u_{l-j+1}\right)\right]\right)(\exp(\mathrm{i}\,\langle q(y),v\rangle)-1)\,\mathbb{P}_{Y|X}(x,\mathrm{d}y)\right|\\
&\le \int \left|\exp\left(\mathrm{i}\left[-g(y)\left(u_1+\ldots+u_{l-j+1}\right)\right]\right)\right|\left|\exp(\mathrm{i}\,\langle q(y),v\rangle)-1\right|\mathbb{P}_{Y|X}(x,\mathrm{d}y)\\
&\le \int |\langle q(y),v\rangle|\,\mathbb{P}_{Y|X}(x,\mathrm{d}y)\\
&= n^{-\frac{1}{2}}\int \left|\left\langle 2(\alpha-y)t+n^{-\frac{1}{2}}t^2,v\right\rangle\right|\mathbb{P}_{Y|X}(x,\mathrm{d}y)\\
&\longrightarrow 0\,,
\end{aligned}$$

as $n\to\infty$. To this end we used that for real $t\in\mathbb{R}$ we have $|\exp(\mathrm{i}\,t)-1|\le|t|$, what becomes clear when thinking of $\exp(\mathrm{i}\,t)$ as a point on the unit circle. Then $|\exp(\mathrm{i}\,t)-1|$ describes the length of a chord and $|t|$ the length of the corresponding arc. Let $\varphi^{\tau-}_{\frac{\alpha+\beta}{2}-Y}$ denotes the characteristic function of $\mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau-,\mathrm{d}y)$. By assumptions (A2) and (B2), and the results in A.5 we find

$$\sum_{j=1}^{l} n\mathbb{E}\left(\mathbb{1}_{X\in U_j}\left(\exp\left(\mathrm{i}\left[-g(Y)\left(u_1+\ldots+u_{l-j+1}\right)+\langle q(Y),v\rangle\right]\right)-1\right)\right)$$
$$\longrightarrow \sum_{j=1}^{l}\left(t_{j-1}-t_j\right)F'_-(\tau)\left(\varphi^{\tau-}_{\frac{\alpha+\beta}{2}-Y}\left(u_1+\ldots+u_{l-j+1}\right)-1\right),$$

as $n\to\infty$, as well as for the third term in Equation (6.11)

$$\sum_{j=1}^{m} n\mathbb{E}\left(\mathbb{1}_{X\in U'_j}\left(\exp\left(\mathrm{i}\left[g(Y)\left(u_{l+1}+\ldots+u_{l+j}\right)+\left\langle q'(Y),w\right\rangle\right]\right)-1\right)\right)$$
$$\longrightarrow \sum_{j=1}^{m}\left(t'_j-t'_{j-1}\right)F'_+(\tau)\left(\varphi^{\tau+}_{Y-\frac{\alpha+\beta}{2}}\left(u_{l+1}+\ldots+u_{l+j}\right)-1\right),$$

as $n\to\infty$. Finally,

$$\varphi_{\left(\pi_T\left(Z_n^{(1)}\right),\pi_T\left(Z_n^{(2)}\right),\pi_T\left(Z_n^{(3)}\right)\right)}(z)=\left(1+\frac{n\left(\left(\varphi_{\left(\pi_T\left(Z_n^{(1)}\right),\pi_T\left(Z_n^{(2)}\right),\pi_T\left(Z_n^{(3)}\right)\right)}(z)\right)^{n^{-1}}-1\right)}{n}\right)^{n}$$
$$\longrightarrow \varphi_{\left(\pi_T(Z^{(1)}),\pi_T(Z^{(2)}),\pi_T(Z^{(3)})\right)}(z),$$

as $n\to\infty$. Tightness of the processes $Z_n^{(1)}$, $Z_n^{(2)}$ and $Z_n^{(3)}$ is already included in the proofs of Corollary 6.11 and Lemma 6.12. □

Lemma 6.14 Suppose that assumptions (A1)-(A3) and (B1)-(B2) hold. For all $d>0$ then

$$\|R_n\|_{[-d,d]^3}\xrightarrow{\mathbb{P}}0,$$

as $n\to\infty$.

Proof. Given $d > 0$, then

$$\begin{aligned}
&\sup_{(t,a,b)\in[-d,d]^3} |R_n(t,a,b)| \\
&\leq 4d\,(|\alpha| \vee |\beta|)\, n^{-\frac{1}{2}} \sup_{t\in[-d,d]} \sum_{i=1}^{n} \left| \mathbb{1}_{X_i \leq \tau + n^{-1}t} - \mathbb{1}_{X_i \leq \tau} \right| \\
&\quad + 4dn^{-\frac{1}{2}} \sup_{t\in[-d,d]} \sum_{i=1}^{n} \left| \mathbb{1}_{X_i \leq \tau + n^{-1}t} - \mathbb{1}_{X_i \leq \tau} \right| |Y_i| + 2d^2 n^{-1} \sup_{t\in[-d,d]} \sum_{i=1}^{n} \left| \mathbb{1}_{X_i \leq \tau + n^{-1}t} - \mathbb{1}_{X_i \leq \tau} \right| \\
&\leq 4d\,(|\alpha| \vee |\beta|)\, n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_{\tau - n^{-1}d < X_i \leq \tau + n^{-1}d} + 4dn^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_{\tau - n^{-1}d < X_i \leq \tau + n^{-1}d} |Y_i| \\
&\qquad + 2d^2 n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\tau - n^{-1}d < X_i \leq \tau + n^{-1}d} \,.
\end{aligned}$$

Lemma 6.10 particularly implies that both, $\sum_{i=1}^{n} \mathbb{1}_{\tau - n^{-1}d < X_i \leq \tau + n^{-1}d}$ and $\sum_{i=1}^{n} \mathbb{1}_{\tau - n^{-1}d < X_i \leq \tau + n^{-1}d} |Y_i|$, converge in distribution. Finally, Slutsky's theorem completes the proof. □

Proof of Theorem 6.7. Given $\gamma \in \mathbb{R}$, define the map $\Phi : D_3 \to D(\mathbb{R}^3)$ with $(f,g,h) \mapsto \gamma f + g + h$ and $\Phi(f,g,h)(x_1,x_2,x_3) = \gamma f(x_1) + g(x_2) + h(x_3)$ for $x = (x_1,x_2,x_3) \in \mathbb{R}^3$. Let us first use Lemma 3.3 and the notations specified therein to show that Φ is continuous. If $(f_n,g_n,h_n)_{n\in\mathbb{N}} \subseteq D_3$ is a sequence with $(f_n,g_n,h_n) \to (f,g,h)$, as $n \to \infty$ with respect to the product topology, then all its projections to $D(\mathbb{R})$ converge and there exist sequences $(\lambda_{k,n})_{n\in\mathbb{N}} \subseteq \Lambda$, $k = 1,2,3$, each of them satisfying condition (i) and (ii) from Lemma 3.3. Set $\lambda_n : \mathbb{R}^3 \to \mathbb{R}^3$, $n \in \mathbb{N}$, with $\lambda_n(x_1,x_2,x_3) = (\lambda_{1,n}(x_1), \lambda_{2,n}(x_2), \lambda_{3,n}(x_3))$ then $(\lambda_n)_{n\in\mathbb{N}} \subseteq \Lambda_3$ and

$$\sup_{x\in\mathbb{R}^3} \|\lambda_n(x) - x\|_\infty \leq \sup_{x_1\in\mathbb{R}} \left|\lambda_{1,n}(x_1) - x_1\right| + \sup_{x_2\in\mathbb{R}} \left|\lambda_{2,n}(x_2) - x_2\right| + \sup_{x_3\in\mathbb{R}} \left|\lambda_{3,n}(x_3) - x_3\right| ,$$

which converges to 0 due to condition (i). Furthermore, for each $a > 0$

$$\begin{aligned}
&\sup_{x\in[-a,a]^3} |\Phi(f_n, g_n, h_n)(\lambda_n(x)) - \Phi(f, g, h)(x)| \\
&= \sup_{(x_1,x_2,x_3)\in[-a,a]^3} \Big|\gamma f_n\left(\lambda_{1,n}(x_1)\right) + g_n\left(\lambda_{2,n}(x_2)\right) \\
&\qquad\qquad + h_n\left(\lambda_{3,n}(x_3)\right) - (\gamma f(x_1) + g(x_2) + h(x_3))\Big| \\
&\le \gamma \sup_{x_1\in[-a,a]} \left|f_n\left(\lambda_{1,n}(x_1)\right) - f(x_1)\right| + \sup_{x_2\in[-a,a]} \left|g_n\left(\lambda_{2,n}(x_2)\right) - g(x_2)\right| \\
&\qquad\qquad + \sup_{x_3\in[-a,a]} \left|h_n\left(\lambda_{3,n}(x_3)\right) - h(x_3)\right| ,
\end{aligned}$$

which also converges to 0, as $n \to \infty$, due to condition (ii). Thus, $\Phi(f_n, g_n, h_n) \to \Phi(f, g, h)$ in $(D(\mathbb{R}^3), s_3)$ and continuity of Φ is proved.

If we set $\gamma = 2(\beta - \alpha)$, then $Z_n = \Phi(Z_n^{(1)}, Z_n^{(2)}, Z_n^{(3)}) + R_n$ and $Z = \Phi(Z^{(1)}, Z^{(2)}, Z^{(3)})$, and by Proposition 6.13 and the CMT, $\Phi\left(Z_n^{(1)}, Z_n^{(2)}, Z_n^{(3)}\right) \xrightarrow{\mathscr{L}} \Phi\left(Z^{(1)}, Z^{(2)}, Z^{(3)}\right)$. By Lemma 6.14 then for each $d > 0$

$$\sup_{(t,a,b)\in[-d,d]^3} \left|Z_n(t, a, b) - \Phi\left(Z_n^{(1)}, Z_n^{(2)}, Z_n^{(3)}\right)(t, a, b)\right| = \sup_{(t,a,b)\in[-d,d]^3} |R_n(t, a, b)| \xrightarrow{\mathbb{P}} 0 ,$$

as $n \to \infty$. If $\lambda_n = \mathrm{Id}_{R^3}$ in Lemma 3.3 then immediately follows $s_3\left(Z_n, \Phi\left(Z_n^{(1)}, Z_n^{(2)}, Z_n^{(3)}\right)\right) \xrightarrow{\mathbb{P}} 0$ and by Slutsky's theorem $Z_n \xrightarrow{\mathscr{L}} Z$ in $D(\mathbb{R}^3)$. □

6.3 Stochastic boundedness of the estimator

Theorem 6.15 Suppose that assumptions **(A1)-(A3)** and **(B1)** hold and $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$, as $n \to \infty$, then

$$\lim_{d\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\left\| \begin{matrix} n(\tau_n - \tau) \\ \sqrt{n}(\alpha_n - \alpha) \\ \sqrt{n}(\beta_n - \beta) \end{matrix} \right\| \ge d \right) = 0 .$$

We introduce the deterministic function Δ_n and adapt the decomposition of Z_n in (6.3) to derive

$$\Delta_n(t, a, b) := \mathbb{E}(Z_n(t, a, b)) = 2(\beta - \alpha)\Delta_n^{(1)}(t) + \Delta_n^{(2)}(a, b) + \Delta_n^{(3)}(t, a, b) , \qquad (6.12)$$

where

$$\begin{aligned}
\Delta_n^{(1)}(t) :=& \mathbb{E}\left(Z_n^{(1)}(t)\right) = n\left[\left(H(\tau+n^{-1}t)-H(\tau)\right)-\frac{\alpha+\beta}{2}\left(F(\tau+n^{-1}t)-F(\tau)\right)\right]\\
\Delta_n^{(2)}(a,b) :=& \mathbb{E}\left(Z_n^{(2)}(a)+Z_n^{(3)}(b)\right) = a^2F(\tau)+b^2\bar{F}(\tau)\\
\Delta_n^{(3)}(t,a,b) :=& \mathbb{E}\left(R_n(t,a,b)\right)\\
=& 2n^{\frac{1}{2}}\left[(b-a)\left(H(\tau+n^{-1}t)-H(\tau)\right)+(a\alpha-b\beta)\left(F(\tau+n^{-1}t)-F(\tau)\right)\right]\\
& +\left(a^2-b^2\right)\left(F(\tau+n^{-1}t)-F(\tau)\right)\\
=& 2n^{\frac{1}{2}}\left[(b-a)\left(\left(H(\tau+n^{-1}t)-H(\tau)\right)-\frac{\alpha+\beta}{2}\left(F(\tau+n^{-1}t)-F(\tau)\right)\right)\right.\\
& \left.+(b+a)\frac{\alpha-\beta}{2}\left(F(\tau+n^{-1}t)-F(\tau)\right)\right]\\
& +\left(a^2-b^2\right)\left(F(\tau+n^{-1}t)-F(\tau)\right). \qquad (6.13)
\end{aligned}$$

Now we will state a lower bound function for Δ_n.

Lemma 6.16 If **(A1)-(A3)** and **(B1)** hold then there exists some $\varepsilon > 0$ and constants $D_1, D_2 > 0$ such that

$$\Delta_n(t,a,b) \geq D_1|t| + D_2(|a| \vee |b|)^2$$

for all (t,a,b) with $\left\|\left(n^{-1}t, n^{-1/2}a, n^{-1/2}b\right)\right\|_\infty \leq \varepsilon$.

Proof. We state the proof for the case when $\alpha < \beta$, the case where $\beta < \alpha$ can be handled in much the same way. Due to assumption **(A2)** and **(B1)** we can formulate the following conditions. First, we find some $\varepsilon_1 > 0$ such that for all $0 < \varepsilon \leq \varepsilon_1$

$$\underline{L}|u-v| \leq |F(u)-F(v)| \leq \overline{L}|u-v|$$

for some constants $0 < \underline{L} < \overline{L} < \infty$ and for all $u, v \in U_\varepsilon(\tau)$. Secondly, define

$$C(\varepsilon) := \left(\frac{\alpha+\beta}{2} - \sup_{x\in(\tau-\varepsilon,\tau)} m(x)\right) \wedge \left(\inf_{x\in(\tau,\tau+\varepsilon)} m(x) - \frac{\alpha+\beta}{2}\right).$$

and note that $C(\varepsilon)$ is monotonically non-decreasing when $\varepsilon \downarrow 0$. From (6.1) we find an $\varepsilon_2 > 0$ sufficiently small such that $C(\varepsilon) > 0$ and

$$\varepsilon\overline{L} < \underline{L}C(\varepsilon)$$

for all $0 < \varepsilon \le \varepsilon_2$. Thirdly, choose $\varepsilon_3 > 0$ such that

$$D_2(\varepsilon) := F(\tau - \varepsilon) \wedge \bar{F}(\tau + \varepsilon) > 0 \text{ and } \varepsilon < \frac{\beta - \alpha}{2}$$

for all $0 < \varepsilon \le \varepsilon_3$. Denote $\varepsilon_0 := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$. Set

$$\begin{aligned} h_n(t) &:= \big(H(\tau + n^{-1}t) - H(\tau)\big) - \frac{\alpha+\beta}{2}\big(F(\tau + n^{-1}t) - F(\tau)\big) \\ &= \begin{cases} \int_{(\tau,\tau+n^{-1}t]} m(x) - \frac{\alpha+\beta}{2} F(\mathrm{d}x) & , t \ge 0 \\ \int_{(\tau+n^{-1}t,\tau]} \frac{\alpha+\beta}{2} - m(x) F(\mathrm{d}x) & , t < 0 \end{cases} \end{aligned}$$

and observe that

$$h_n(t) \ge \underline{L} C(\varepsilon) n^{-1} |t| \ge 0 \tag{6.14}$$

for all $n^{-1}t \in (-\varepsilon_0, \varepsilon_0)$. Furthermore, set

$$\begin{aligned} g_n(t) &:= \frac{\alpha-\beta}{2}\big(F(\tau + n^{-1}t) - F(\tau)\big) \\ &= \begin{cases} \frac{\alpha-\beta}{2} \int_{(\tau,\tau+n^{-1}t]} F(\mathrm{d}x) & , t \ge 0 \\ \frac{\beta-\alpha}{2} \int_{(\tau+n^{-1}t,\tau]} F(\mathrm{d}x) & , t < 0 \end{cases} \end{aligned}$$

and find that $g_n(t) < 0$ for all $t > 0$, and $g_n(t) > 0$ for all $t < 0$. It is also true, that

$$|g_n(t)| \le \frac{\beta - \alpha}{2} \overline{L} n^{-1} |t| \tag{6.15}$$

for all $n^{-1}t \in (-\varepsilon_0, \varepsilon_0)$. For all (t, a, b) with $\left\|\big(n^{-1}t, n^{-1/2}a, n^{-1/2}b\big)\right\|_\infty \le \varepsilon_0$ we can write

$$\begin{aligned} &\Delta_n(t, a, b) \\ &= 2(\beta-\alpha) n h_n(t) + 2n^{\frac{1}{2}} \left[(b-a)h_n(t) + (b+a)g_n(t)\right] + a^2 F(\tau + n^{-1}t) + b^2 \bar{F}(\tau + n^{-1}t) \\ &\ge 2(\beta-\alpha) n h_n(t) + 2n^{\frac{1}{2}} \min_{|a| \vee |b| \le n^{\frac{1}{2}} \varepsilon_0} \left[(b-a)h_n(t) + (b+a)g_n(t)\right] + D_2(|a| \vee |b|)^2 \,. \end{aligned}$$

For some fixed $t \in \mathbb{R}$ the term in square brackets, $(b-a)h_n(t) + (b+a)g_n(t)$, is a real valued, linear function in a and b, where the feasible region is stated by the linear inequality constraints $|a| \vee |b| \le n^{1/2}\varepsilon_0$. To find the minimum on this set is a linear programming problem. It is known that the minimum value is always attained on the vertices of this region (cf. [Sie96, Theorem 2.2.2]), in this case on the set $A := \left\{(-n^{1/2}\varepsilon_0, n^{1/2}\varepsilon_0), (n^{1/2}\varepsilon_0, n^{1/2}\varepsilon_0), (n^{1/2}\varepsilon_0, -n^{1/2}\varepsilon_0), (-n^{1/2}\varepsilon_0, -n^{1/2}\varepsilon_0)\right\}$. Thus we continue

$$
\begin{aligned}
&\Delta_n(t,a,b)\\
&\quad \geq \quad 2(\beta-\alpha)nh_n(t) + 2n^{\frac{1}{2}} \min_{(a,b)\in A} \left[(b-a)h_n(t) + (b+a)g_n(t)\right] + D_2(|a| \vee |b|)^2\\
&\quad \geq \quad 2(\beta-\alpha)nh_n(t) + 4n\varepsilon_0 \min\left\{-|g_n(t)|, -h_n(t)\right\} + D_2(|a| \vee |b|)^2\\
&\quad = \quad \min\left\{2(\beta-\alpha)nh_n(t) - 4n\varepsilon_0|g_n(t)|, 2(\beta-\alpha)nh_n(t) - 4n\varepsilon_0 h_n(t)\right\}\\
&\qquad\qquad + D_2(|a| \vee |b|)^2\\
&\quad \overset{(6.14),(6.15)}{\geq} \quad \min\left\{2(\beta-\alpha)\left[\underline{L}C(\varepsilon_0) - \varepsilon_0\overline{L}\right], \left[2(\beta-\alpha) - 4\varepsilon_0\right]\underline{L}C(\varepsilon_0)\right\}|t|\\
&\qquad\qquad + D_2(|a| \vee |b|)^2\\
&\quad = \quad D_1|t| + D_2(|a| \vee |b|)^2 ,
\end{aligned}
$$

where $D_1 := \min\left\{\left[2(\beta-\alpha) - 4\varepsilon_0\right]\underline{L}C(\varepsilon_0), 2(\beta-\alpha)\left[\underline{L}C(\varepsilon_0) - \varepsilon_0\overline{L}\right]\right\} > 0$. □

Proof of Theorem 6.15. Let $d > 0$ and choose $\varepsilon > 0$ being admissible in the sense of Lemma 6.16. With the appropriate positive constants D_1 and D_2 from Lemma 6.16 define

$$
\Delta(t,a,b) = D_1|t| + D_2\left(|a| \vee |b|\right)^2 . \tag{6.16}
$$

Within this proof for brevity use D_1 and D_2 and a constant $D > 0$ as generic constants (cf. Remark 4.1). Write $y_n := \left(n(\tau_n - \tau), \sqrt{n}(\alpha_n - \alpha), \sqrt{n}(\beta_n - \beta)\right)^\top$, define the diagonal matrix

$$
M := \operatorname{diag}(n^{-1}, n^{-\frac{1}{2}}, n^{-\frac{1}{2}})
$$

and note that

$$
\mathbb{P}\left(\left\| \begin{matrix} n(\tau_n - \tau) \\ \sqrt{n}(\alpha_n - \alpha) \\ \sqrt{n}(\beta_n - \beta) \end{matrix} \right\| \geq d\right) = \mathbb{P}\left(\|y_n\| \geq d\right) \leq \mathbb{P}\left(\|y_n\| \geq d\ , \|My_n\| \leq \varepsilon\right) + \mathbb{P}\left(\|My_n\| > \varepsilon\right) . \tag{6.17}
$$

By $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$ we already know, that $\lim\limits_{n\to\infty} \mathbb{P}\left(\|My_n\| > \varepsilon\right) = 0$. Thus, it only remains to take account of the first term on the right hand side of (6.17). In the following we choose the maximum norm $\|\cdot\|_\infty$ for $\|\cdot\|$ while knowing that in finite-dimensional vector spaces all norms are equivalent. Having in mind that $(\tau_n, \alpha_n, \beta_n) \in \operatorname{Argmin}(S_n)$, observe that $y_n \in \operatorname{Argmin}(Z_n)$ by Lemma 3.9. Set

$$J := \{u = (t, a, b);\ d \leq \|u\|_\infty \text{ and } \|Mu\|_\infty \leq \varepsilon\}\ .$$

If $\|y_n\|_\infty \geq d$ and $\|My_n\|_\infty \leq \varepsilon$ then there exists some $u = (t, a, b) \in J$ such that

$$S_n(\tau_n + n^{-1}t, \alpha + n^{-\frac{1}{2}}a, \beta + n^{-\frac{1}{2}}b) \leq S_n(\tau, \alpha, \beta)\ .$$

By the definition of the rescaled process Z_n in (6.2), an application of Lemma 6.16, Equation (6.16) and the decomposition of Δ_n in (6.12) then

$$\begin{aligned}
&\mathbb{P}\big(\|y_n\|_\infty \geq d\ , \|My_n\|_\infty \leq \varepsilon\big) \\
&\leq \mathbb{P}\Big(\bigcup_{u\in J}\{Z_n(u) \leq 0\}\Big) \\
&= \mathbb{P}\Big(\bigcup_{u\in J}\{\Delta_n(u) - Z_n(u) \geq \Delta_n(u)\}\Big) \\
&\leq \mathbb{P}\Big(\bigcup_{(t,a,b)\in J}\{\Delta_n(t,a,b) - Z_n(t,a,b) \geq \Delta(t,a,b)\}\Big) \\
&\leq \mathbb{P}\Big(\bigcup_{(t,a,b)\in J}\{|\Delta_n(t,a,b) - Z_n(t,a,b)| \geq \Delta(t,a,b)\}\Big) \\
&\leq \mathbb{P}\Big(\bigcup_{(t,a,b)\in J}\Big\{\big|2(\alpha-\beta)\big(Z_n^{(1)}(t) - \Delta_n^{(1)}(t)\big)\big| \geq \frac{1}{3}D\Delta(t,a,b)\Big\}\Big) \\
&\qquad + \mathbb{P}\Big(\bigcup_{(t,a,b)\in J}\Big\{\big|\Delta_n^{(2)}(a,b) - \big(Z_n^{(2)}(a) + Z_n^{(3)}(b)\big)\big| \geq \frac{1}{3}D\Delta(t,a,b)\Big\}\Big) \\
&\qquad\quad + \mathbb{P}\Big(\bigcup_{(t,a,b)\in J}\Big\{\big|\Delta_n^{(3)}(t,a,b) - R_n(t,a,b)\big| \geq \frac{1}{3}D\Delta(t,a,b)\Big\}\Big)\ . \qquad (6.18)
\end{aligned}$$

Now split J into $J_1 := \{u = (t,a,b) \in J; |t| \geq d\}$ and $J_2 := \{u = (t,a,b) \in J\ ; |t| < d\}$. For every $(t,a,b) \in J_1$ then by Lemma 6.16, $\Delta(t,a,b) \geq D_1|t|$. If $(t,a,b) \in J_2$ then $|t| < d \leq |a| \vee |b| \leq n^{1/2}\varepsilon$ and, hence, $\Delta(t,a,b) \geq D_2 d^2$. For the first term on the right hand side of (6.18) use (6.3) and recall the notion of W_n in (4.5) to see that

$$Z_n^{(1)}(t) - \Delta_n^{(1)}(t) = n\big(W_n(\tau + n^{-1}t) - W_n(\tau)\big)\ .$$

It follows that

$$
\begin{aligned}
&\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|2(\alpha-\beta)\left(Z_n^{(1)}(t)-\Delta_n^{(1)}(t)\right)\right|\geq D\Delta(t,a,b)\right\}\right)\\
&\leq \mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{n\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D\Delta(t,a,b)\right\}\right)\\
&\leq \mathbb{P}\left(\bigcup_{(t,a,b)\in J_1}\left\{n\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D\Delta(t,a,b)\right\}\right)\\
&\qquad\qquad+\mathbb{P}\left(\bigcup_{(t,a,b)\in J_2}\left\{n\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D\Delta(t,a,b)\right\}\right)\\
&\leq \mathbb{P}\left(\bigcup_{d\leq|t|\leq n\varepsilon}\left\{n\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D_1|t|\right\}\right)\\
&\qquad\qquad+\mathbb{P}\left(\bigcup_{|t|<d}\left\{n\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D_2d^2\right\}\right). \qquad (6.19)
\end{aligned}
$$

We apply Lemmas 4.17 and 4.18, while choosing $\nu=1$ and $\gamma=2$, to get

$$
\lim_{d\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|2(\alpha-\beta)\left(Z_n^{(1)}(t)-\Delta_n^{(1)}(t)\right)\right|\geq D\Delta(t,a,b)\right\}\right)=0\,.
$$

Now turn to the second probability in (6.18). With the estimations

$$
\left|a^2\left(F(\tau)-F_n(\tau)\right)+b^2\left(\bar{F}(\tau)-\bar{F}_n(\tau)\right)\right|\leq 2\left(|a|\vee|b|\right)^2\left|F_n(\tau)-F(\tau)\right|
$$

and

$$
\begin{aligned}
&2n^{-\frac{1}{2}}\left|a\sum_{i=1}^n\mathbb{1}_{\{X_i\leq\tau\}}(\alpha-Y_i)+b\sum_{i=1}^n\mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)\right|\\
&\leq 2n^{-\frac{1}{2}}\left(|a|\left|\sum_{i=1}^n\mathbb{1}_{\{X_i\leq\tau\}}(\alpha-Y_i)\right|+|b|\left|\sum_{i=1}^n\mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)\right|\right)\\
&\leq 2n^{-\frac{1}{2}}\left(|a|\vee|b|\right)\left(\left|\sum_{i=1}^n\mathbb{1}_{\{X_i\leq\tau\}}(\alpha-Y_i)\right|+\left|\sum_{i=1}^n\mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)\right|\right),
\end{aligned}
$$

we conclude

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|\Delta_n^{(2)}(a,b)-\left(Z_n^{(2)}(a)+Z_n^{(3)}(b)\right)\right|\geq D\Delta(t,a,b)\right\}\right)\\
&\overset{(6.3)}{=}\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|a^2\left(F(\tau)-F_n(\tau)\right)+b^2\left(\bar{F}(\tau)-\bar{F}_n(\tau)\right)\right.\right.\right.\\
&\qquad\qquad\left.\left.\left.-2n^{-\frac{1}{2}}\left[a\sum_{i=1}^n \mathbb{1}_{\{X_i\leq\tau\}}(\alpha-Y_i)+b\sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)\right]\right|\geq D\Delta(t,a,b)\right\}\right)\\
&\leq\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{(|a|\vee|b|)^2\left|F_n(\tau)-F(\tau)\right|\right.\right.\\
&\quad\left.\left.+n^{-\frac{1}{2}}(|a|\vee|b|)\left(\left|\sum_{i=1}^n \mathbb{1}_{\{X_i\leq\tau\}}(\alpha-Y_i)\right|+\left|\sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)\right|\right)\geq D\Delta(t,a,b)\right\}\right).
\end{aligned}$$

Now define the disjoint sets $\tilde{J}_1 := \{u\in J; |a|\vee|b|\geq d\}$ and $\tilde{J}_2 := \{u\in J; |a|\vee|b|< d\}$, and split the probability by $\tilde{J}_1$ and $\tilde{J}_2$ into $\mathbb{P}(A_n^{(1)})$ and $\mathbb{P}(A_n^{(2)})$. For all $u\in\tilde{J}_1$ then $\Delta(t,a,b)\geq D_2(|a|\vee|b|)^2$ and we estimate

$$\begin{aligned}
&\mathbb{P}\left(A_n^{(1)}\right)\\
&=\mathbb{P}\left(\bigcup_{(t,a,b)\in\tilde{J}_1}\left\{(|a|\vee|b|)^2\left|F_n(\tau)-F(\tau)\right|\right.\right.\\
&\quad\left.\left.+n^{-\frac{1}{2}}(|a|\vee|b|)\left(\left|\sum_{i=1}^n \mathbb{1}_{\{X_i\leq\tau\}}(\alpha-Y_i)\right|+\left|\sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)\right|\right)\geq D_2(|a|\vee|b|)^2\right\}\right)\\
&\leq\mathbb{P}\left(\left|F_n(\tau)-F(\tau)\right|\geq D_2\right)+\mathbb{P}\left(n^{-\frac{1}{2}}\left|\sum_{i=1}^n \mathbb{1}_{\{X_i\leq\tau\}}(\alpha-Y_i)\right|\geq D_2 d\right)\\
&\qquad\qquad+\mathbb{P}\left(n^{-\frac{1}{2}}\left|\sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}}(\beta-Y_i)\right|\geq D_2 d\right). \qquad (6.20)
\end{aligned}$$

The first term of (6.20) converges to 0 as $n\to\infty$, since $F_n(\tau)$ converges to $F(\tau)$, $\mathbb{P}$-almost surely. As a consequence of the Central Limit Theorem (CLT)

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le \tau\}} (\alpha - Y_i) \xrightarrow{\mathscr{L}} \mathcal{N}(0, T(\tau)) \,, \tag{6.21}$$

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \tau\}} (\beta - Y_i) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \bar{T}(\tau)) \,, \tag{6.22}$$

as $n \to \infty$. For a random variable $W \sim \mathcal{N}(0,1)$ we can write

$$\begin{aligned}
&\mathbb{P}\left(n^{-\frac{1}{2}} \left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le \tau\}} (\alpha - Y_i) \right| \ge D_2 d \right) \\
&= \left[\mathbb{P}\left(n^{-\frac{1}{2}} \left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le \tau\}} (\alpha - Y_i) \right| \ge D_2 d \right) - \mathbb{P}\left(\sqrt{T(\tau)} |W| \ge D_2 d \right) \right] \\
&\qquad + \mathbb{P}\left(|W| \ge \frac{D_2}{\sqrt{T(\tau)}} d \right) .
\end{aligned}$$

From (6.21) we conclude that the term in square brackets converges to zero as $n \to \infty$ for all $d > 0$. Using Markov's inequality for the second term leads to

$$\mathbb{P}\left(|W| \ge \frac{D_2}{\sqrt{T(\tau)}} d \right) \le \frac{\sqrt{T(\tau)} \mathbb{E}\,|W|}{D_2} d^{-1} = \frac{\sqrt{\frac{2}{\pi} T(\tau)}}{D_2} d^{-1} .$$

After applying similar arguments to the last term in (6.20), we find

$$\limsup_{n\to\infty} \mathbb{P}(A_n^{(1)}) \le \frac{\sqrt{\frac{2}{\pi}} \left(\sqrt{T(\tau)} + \sqrt{\bar{T}(\tau)} \right)}{D_2} d^{-1} .$$

If $u \in \tilde{J}_2$ then $d \le |t| \le n\varepsilon$ and, thus, $\Delta(t, a, b) \ge D_1 d + D_2 (|a| \vee |b|)^2$. Then we estimate

$$\begin{aligned}
&\mathbb{P}\left(A_n^{(2)} \right) \\
&\le \mathbb{P}\Bigg(\bigcup_{(t,a,b) \in \tilde{J}_2} \Bigg\{ (|a| \vee |b|)^2 \, |F_n(\tau) - F(\tau)| \\
&\qquad + n^{-\frac{1}{2}} (|a| \vee |b|) \left(\left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le \tau\}} (\alpha - Y_i) \right| + \left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \tau\}} (\beta - Y_i) \right| \right) \\
&\qquad \ge D_1 d + D_2 (|a| \vee |b|)^2 \Bigg\} \Bigg)
\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{P}\Bigg(\bigcup_{(t,a,b)\in \tilde{J}_2} \big\{ (|a|\vee|b|)^2 \left(|F_n(\tau)-F(\tau)|-D_2\right) \ge D_1 d \big\} \Bigg) \\
&\qquad + \mathbb{P}\Bigg(\bigcup_{(t,a,b)\in \tilde{J}_2} \Bigg\{ n^{-\frac{1}{2}} \left(|a|\vee|b|\right) \left(\left| \sum_{i=1}^n \mathbb{1}_{\{X_i\le\tau\}} (\alpha - Y_i) \right| + \left| \sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}} (\beta - Y_i) \right| \right) \\
&\qquad\qquad \ge D_1 d + D_2 \left(|a|\vee|b|\right)^2 \Bigg\} \Bigg) \\
&\le \mathbb{P}\left(|F_n(\tau)-F(\tau)| > D_2\right) \\
&\qquad + \mathbb{P}\left(n^{-1} \left(\left| \sum_{i=1}^n \mathbb{1}_{\{X_i\le\tau\}} (\alpha - Y_i) \right| + \left| \sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}} (\beta - Y_i) \right| \right)^2 \ge 4 D_2 D_1 d \right) .
\end{aligned}
$$

To establish the last inequality, regard the events as inequalities of a quadratic polynomial in $|a|\vee|b|$ and use its discriminant to deduce solutions. As above the first term converges to 0 as $n\to\infty$. If $W,\tilde{W}\sim\mathcal{N}(0,1)$, then

$$
\begin{aligned}
&\mathbb{P}\left(n^{-1} \left(\left| \sum_{i=1}^n \mathbb{1}_{\{X_i\le\tau\}} (\alpha - Y_i) \right| + \left| \sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}} (\beta - Y_i) \right| \right)^2 \ge D_2 D_1 d \right) \\
&= \Bigg[\mathbb{P}\left(n^{-1} \left(\left| \sum_{i=1}^n \mathbb{1}_{\{X_i\le\tau\}} (\alpha - Y_i) \right| + \left| \sum_{i=1}^n \mathbb{1}_{\{X_i>\tau\}} (\beta - Y_i) \right| \right)^2 \ge D_2 D_1 d \right) \\
&\qquad - \mathbb{P}\left(\left(\left|\sqrt{T(\tau)}W\right| + \left|\sqrt{\bar{T}(\tau)}\tilde{W}\right| \right)^2 \ge D_2 D_1 d \right) \Bigg] \\
&\qquad + \mathbb{P}\left(\left(\left|\sqrt{T(\tau)}W\right| + \left|\sqrt{\bar{T}(\tau)}\tilde{W}\right| \right)^2 \ge D_2 D_1 d \right) ,
\end{aligned}
$$

where the term in square brackets converges to zero as $n\to\infty$ by (6.21) and (6.22), and the CMT. Applying Markov's inequality yields

$$
\mathbb{P}\left(\left(\left|\sqrt{T(\tau)}W\right| + \left|\sqrt{\bar{T}(\tau)}\tilde{W}\right| \right)^2 \ge D_2 D_1 d \right) \le \frac{\mathbb{E}\left(\left|\sqrt{T(\tau)}W\right| + \left|\sqrt{\bar{T}(\tau)}\tilde{W}\right| \right)^2}{D_2 D_1} d^{-1} ,
$$

and, thus,

$$
\limsup_{n\to\infty} \mathbb{P}(A_n^{(2)}) \le \frac{\mathbb{E}\left(\left|\sqrt{T(\tau)}W\right| + \left|\sqrt{\bar{T}(\tau)}\tilde{W}\right| \right)^2}{D_2 D_1} d^{-1} .
$$

Finally,

$$\lim_{d\to\infty}\lim_{n\to\infty}\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|\Delta_n^{(2)}(a,b)-\left(Z_n^{(2)}(a)+Z_n^{(3)}(b)\right)\right|\geq D\Delta(t,a,b)\right\}\right)=0\,.$$

We consider the last term on the right hand side of Inequality (6.18). Use the fact that

$$(a\alpha-b\beta)+(b-a)Y_i=(b-a)\left(Y_i-\frac{\alpha+\beta}{2}\right)+\frac{1}{2}(a+b)(\alpha-\beta)$$

and by using (6.3) and (6.13) derive

$$\begin{aligned}
&R_n(t,a,b)-\Delta_n^{(3)}(t,a,b)\\
&=2(b-a)n^{-\frac{1}{2}}\sum_{i=1}^{n}\Bigg[\left(\mathbb{1}_{X_i\leq\tau+n^{-1}t}-\mathbb{1}_{X_i\leq\tau}\right)\left(Y_i-\frac{\alpha+\beta}{2}\right)\\
&\qquad\qquad-\mathbb{E}\left(\left(\mathbb{1}_{X_i\leq\tau+n^{-1}t}-\mathbb{1}_{X_i\leq\tau}\right)\left(Y_i-\frac{\alpha+\beta}{2}\right)\right)\Bigg]\\
&\qquad+(a+b)(\alpha-\beta)n^{-\frac{1}{2}}\sum_{i=1}^{n}\left[\left(\mathbb{1}_{X_i\leq\tau+n^{-1}t}-\mathbb{1}_{X_i\leq\tau}\right)-\mathbb{E}\left(\mathbb{1}_{X_i\leq\tau+n^{-1}t}-\mathbb{1}_{X_i\leq\tau}\right)\right]\\
&\qquad+(a^2-b^2)n^{-1}\sum_{i=1}^{n}\left[\left(\mathbb{1}_{X_i\leq\tau+n^{-1}t}-\mathbb{1}_{X_i\leq\tau}\right)-\mathbb{E}\left(\mathbb{1}_{X_i\leq\tau+n^{-1}t}-\mathbb{1}_{X_i\leq\tau}\right)\right]\\
&=:C_n^{(1)}(t,a,b)+C_n^{(2)}(t,a,b)+C_n^{(3)}(t,a,b)\,.
\end{aligned}$$

and, hence,

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|\Delta_n^{(3)}(t,a,b)-R_n(t,a,b)\right|\geq D\Delta(t,a,b)\right\}\right)\\
&\leq\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(1)}(t,a,b)\right|\geq\frac{1}{3}D\Delta(t,a,b)\right\}\right)\\
&\qquad+\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(2)}(t,a,b)\right|\geq\frac{1}{3}D\Delta(t,a,b)\right\}\right)\\
&\qquad+\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(3)}(t,a,b)\right|\geq\frac{1}{3}D\Delta(t,a,b)\right\}\right)\,. \qquad (6.23)
\end{aligned}$$

Again we split J into $J_1=\{u\in J;|t|\geq d\}$ and $J_2=\{u\in J;|t|<d\}$. By recalling the notion of the process W_n in (4.5) and solving an inequality of a quadratic polynomial in $(|a|\vee|b|)$

we obtain for the first term in (6.23)

$$
\begin{aligned}
&\mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(1)}(t,a,b)\right|\geq D\Delta(t,a,b)\right\}\Bigg)\\
&\leq \mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\left\{4(|a|\vee|b|)n^{\frac{1}{2}}\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D_1|t|+D_2(|a|\vee|b|)^2\right\}\Bigg)\\
&\leq \mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J_1}\left\{(|a|\vee|b|)n^{\frac{1}{2}}\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D_1|t|+D_2(|a|\vee|b|)^2\right\}\Bigg)\\
&\qquad\qquad+\mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J_2}\left\{n^{\frac{1}{2}}\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D_2 d\right\}\Bigg)\\
&\leq \mathbb{P}\Bigg(\bigcup_{d<|t|\leq n\varepsilon}\left\{n\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|^2\geq D_1D_2|t|\right\}\Bigg)\\
&\qquad\qquad+\mathbb{P}\Bigg(\bigcup_{|t|\leq d}\left\{n^{\frac{1}{2}}\left|W_n(\tau+n^{-1}t)-W_n(\tau)\right|\geq D_2 d\right\}\Bigg)\,.
\end{aligned}
\tag{6.24}
$$

By choosing $\nu=1$, $\lambda=1/2$ and $\gamma=1$, we obtain by Lemmas 4.17 and 4.18,

$$
\limsup_{n\to\infty}\mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(1)}(t,a,b)\right|\geq D\Delta(t,a,b)\right\}\Bigg)=0\,, \tag{6.25}
$$

for all $d>0$. Now consider the second term in the right hand side of (6.23) and observe that $C_n^{(2)}(t,a,b)=(a+b)(\alpha-\beta)n^{1/2}V_{1,n}^{(2)}(t)$, where $V_{1,n}^{(2)}$ is the process occurring in the decomposition of W_n in (4.25). The corresponding statements for the process $V_{1,n}^{(2)}$ are contained in the proofs of Lemmas 4.17 and 4.18. Thus

$$
\begin{aligned}
&\mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(2)}(t,a,b)\right|\geq D\Delta(t,a,b)\right\}\Bigg)\\
&\leq \mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\left\{2\left(|\alpha|+|\beta|\right)\left(|a|\vee|b|\right)n^{\frac{1}{2}}\left|V_{1,n}^{(2)}(t)\right|\geq D_1|t|+D_2(|a|\vee|b|)^2\right\}\Bigg)
\end{aligned}
$$

and, what was already shown in the discussion about the process $C_n^{(1)}$,

$$
\limsup_{n\to\infty}\mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(2)}(t,a,b)\right|\geq D\Delta(t,a,b)\right\}\Bigg)=0\,,
$$

for all $d > 0$. Since $C_n^{(3)}(t,a,b) = (a^2 - b^2)V_{1,n}^{(2)}(t)$, it follows

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(3)}(t,a,b)\right| \geq D\Delta(t,a,b)\right\}\right)\\
&= \mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{(|a|\vee|b|)^2\left|V_{1,n}^{(2)}(t)\right| \geq D_1|t| + D_2(|a|\vee|b|)^2\right\}\right)\\
&\leq \mathbb{P}\left(\bigcup_{(t,a,b)\in J_1}\left\{(|a|\vee|b|)^2\left|V_{1,n}^{(2)}(t)\right| \geq D_1|t| + D_2(|a|\vee|b|)^2\right\}\right)\\
&\qquad\qquad + \mathbb{P}\left(\bigcup_{(t,a,b)\in J_2}\left\{(|a|\vee|b|)^2\left|V_{1,n}^{(2)}(t)\right| \geq D_2(|a|\vee|b|)^2\right\}\right)\\
&\leq \mathbb{P}\left(\bigcup_{d<|t|\leq n\varepsilon}\left\{\varepsilon^2 n\left|V_{1,n}^{(2)}(t)\right| \geq D_1|t|\right\}\right) + \mathbb{P}\left(\bigcup_{|t|\leq d}\left\{\varepsilon^2 n\left|V_{1,n}^{(2)}(t)\right| \geq D_2 d^2\right\}\right).
\end{aligned}$$

Applying Lemma 4.17 and 4.18 while choosing $\nu = 1$, $\lambda = 1$ and $\gamma = 2$ gives

$$\limsup_{n\to\infty}\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|C_n^{(3)}(t,a,b)\right| \geq D\Delta(t,a,b)\right\}\right) \leq D\left(d^{-1} + d^{-3}\right)$$

and, therefore,

$$\lim_{d\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|\Delta_n^{(3)}(t,a,b) - R_n(t,a,b)\right| \geq D\Delta(t,a,b)\right\}\right) = 0\,. \qquad \square$$

6.4 Distributional convergence and asymptotic confidence region

Being equipped with the results of the last sections we are able to proof the main statement of this chapter.

Theorem 6.17 If (A1)-(A3) and (B1)-(B2) hold and $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$, as $n \to \infty$, then

$$\limsup_{n\to\infty}\mathbb{P}\left(\left(n(\tau_n - \tau), n^{\frac{1}{2}}(\alpha_n - \alpha), n^{\frac{1}{2}}(\beta_n - \beta)\right) \in C\right) \leq \mathbb{P}(\mathrm{Arginf}(Z) \cap C \neq \emptyset)$$

for all closed $C \subseteq \mathbb{R}^3$ and

$$\begin{aligned}&\operatorname{Arginf}(Z)\\&=\operatorname*{Arginf}_{t\in\mathbb{R}}\big(2(\beta-\alpha)Z^{(1)}(t)\big)\times\left\{\frac{\sqrt{\mathbb{E}\left(\mathbb{1}_{X\le\tau}(\alpha-Y)^2\right)}}{F(\tau)}W\right\}\times\left\{\frac{\sqrt{\mathbb{E}\left(\mathbb{1}_{X>\tau}(\beta-Y)^2\right)}}{\bar{F}(\tau)}W'\right\},\end{aligned} \tag{6.26}$$

where W and W' are two independent standard normal distributed random variables. Furthermore, $Z^{(1)}$ is a two-sided compound Poisson process independent of W and W' and with intensities $F'_+(\tau)$ for $t \ge 0$ and $F'_-(\tau)$ for $t < 0$, and jump size distributions $\mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\cdot)$ and $\mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau-,\cdot)$, respectively.

Lemma 6.18 Suppose that (A1)-(A3) and (B1)-(B2) hold and Z is the process defined in Definition 6.5. Then the set $\operatorname{Arginf}(Z)$ is nonempty and compact and Equation (6.26) is valid.

Proof. Using the calculation rules for infima of sums in [RW98, Excercise 1.36], we get

$$\inf_{(t,a,b)\in\mathbb{R}^3} Z(t,a,b) = \inf_{t\in\mathbb{R}} 2(\beta-\alpha)Z^{(1)}(t) + \inf_{a\in\mathbb{R}} Z^{(2)}(a) + \inf_{b\in\mathbb{R}} Z^{(3)}(b)$$

and, therefore, we can write the Arginf-set as the cartesian product

$$\operatorname*{Arginf}_{(t,a,b)\in R^3} Z(t,a,b) = \operatorname*{Arginf}_{t\in\mathbb{R}}\big(2(\beta-\alpha)Z^{(1)}(t)\big) \times \operatorname*{Arginf}_{a\in\mathbb{R}} Z^{(2)}(a) \times \operatorname*{Arginf}_{b\in\mathbb{R}} Z^{(3)}(b)\,.$$

By Tychonov's theorem it is sufficient to show that each of these sets is compact. Let $(N_1(t))_{t\ge 0}$ be a Poisson process with intensity $F'_+(\tau)$ and let $(\xi'_i)_{i\in\mathbb{N}}$ be i. i. d. copies of $\xi' \sim \mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\cdot)$ which are independent of N_1. For all $t \ge 0$ then

$$2(\beta-\alpha)Z^{(1)}(t) = \sum_{i=1}^{N_1(t)} 2(\beta-\alpha)\xi'_i\,.$$

Since $N_1(t) \to \infty$ almost surely, as $t \to \infty$, an application of the SLLN gives

$$\frac{1}{N_1(t)}\sum_{i=1}^{N_1(t)} 2(\beta-\alpha)\xi'_i \longrightarrow 2(\beta-\alpha)\mathbb{E}(\xi')$$

almost surely, as $t \to \infty$. By Lemma 6.6 we find that $2(\beta-\alpha)\mathbb{E}(\xi') > 0$ and, thus, $2(\beta-\alpha)Z^{(1)}(t) \to \infty$ almost surely, as $t \to \infty$. Repeating the previous arguments to a

$F'_{-}(\tau)$-rate Poisson process $(N_2(-t))_{t<0}$ and $\xi \sim \mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau+,\cdot)$ leads to $2(\beta-\alpha)Z^{(1)}(t) \to \infty$ almost surely, as $t \to -\infty$. If

$$\bar{Z}^{(1)}(t) := \min\left\{\gamma \in \bar{\mathbb{R}}; \exists x_n \to x \text{ with } Z^{(1)}(x_n) \to \gamma\right\}$$

is the lower-semicontinuous regularization of $Z^{(1)}$, then

$$\operatorname*{Arginf}_{t\in\mathbb{R}}\left(2(\beta-\alpha)Z^{(1)}(t)\right) = \operatorname*{Argmin}_{t\in\mathbb{R}}\left(2(\beta-\alpha)\bar{Z}^{(1)}(t)\right),$$

see [Fer15, Lemma 2.2]. By [RW98, Theorem 1.9] then $\operatorname{Arginf}_{t\in\mathbb{R}}(2(\beta-\alpha)Z^{(1)}(t))$ is nonempty and compact. Now consider the set

$$\operatorname*{Arginf}_{a\in\mathbb{R}}\left(Z^{(2)}(a)\right) = \operatorname*{Arginf}_{a\in\mathbb{R}}\left(2\sqrt{\mathbb{E}\left(\mathbb{1}_{X\le\tau}(\alpha-Y)^2\right)}Wa + F(\tau)a^2\right)$$

By merely minimizing a quadratic function, we immediately specify the one-element set

$$\operatorname*{Arginf}_{a\in\mathbb{R}} Z^{(2)}(a) = \left\{-\frac{\sqrt{\mathbb{E}\left(\mathbb{1}_{X\le\tau}(\alpha-Y)^2\right)}}{F(\tau)}W\right\} \sim \left\{\frac{\sqrt{\mathbb{E}\left(\mathbb{1}_{X\le\tau}(\alpha-Y)^2\right)}}{F(\tau)}W\right\},$$

which is compact. Finally, by similar arguments for the process $(Z^{(3)}(b))_{b\in\mathbb{R}}$ we find that

$$\operatorname*{Arginf}_{b\in\mathbb{R}} Z^{(3)}(b) = \left\{-\frac{\sqrt{\mathbb{E}\left(\mathbb{1}_{X>\tau}(\beta-Y)^2\right)}}{\bar{F}(\tau)}W'\right\} \sim \left\{\frac{\sqrt{\mathbb{E}\left(\mathbb{1}_{X>\tau}(\beta-Y)^2\right)}}{\bar{F}(\tau)}W'\right\}. \qquad \square$$

Proof of Theorem 6.17. By Lemma 3.9, $(n(\tau_n-\tau), n^{1/2}(\alpha_n-\alpha), n^{1/2}(\beta_n-\beta)) \in \operatorname{Arginf}(Z_n)$ and $Z_n \in D(\mathbb{R}^3)$ for each $n \in \mathbb{N}$. Applying Proposition 3.12 together with the results of Theorems 6.7, 6.15 and 6.18 prove the theorem. $\square$

Corollary 6.19 (Confidence region) Under the assumptions of Theorem 6.17, let G be the distribution function of the random variable $\gamma := \sup_{x\in\operatorname{Arginf}(Z)} \|x\|_\infty$, and q_η be the $(1-\eta)$-quantile of G, i.e. $q_\eta = G^{-1}(1-\eta)$. Then

$$I_n(r) := \left(\tau_n - rn^{-1}, \tau_n + rn^{-1}\right) \times \left(\alpha_n - rn^{-\frac{1}{2}}, \alpha_n + rn^{-\frac{1}{2}}\right) \times \left(\beta_n - rn^{-\frac{1}{2}}, \beta_n + rn^{-\frac{1}{2}}\right)$$

is an asymptotic confidence region for (τ, α, β) at level $1-\eta$ for all $r > q_\eta$.

Proof. The result can be derived by explanations made in section 4 in [Fer15]. Let $r > q_\eta$. By Lemma 6.18 $\operatorname{Arginf}(Z) \subseteq \mathbb{R}^3$ is compact. Additionally, the maximum-norm on

R^d is continuous and we can use Theorem 6.17 and Proposition 3.14 to conclude

$$\limsup_{n\to\infty} \mathbb{P}\left(\left\|\left(n(\tau_n-\tau), n^{\frac{1}{2}}(\alpha_n-\alpha), n^{\frac{1}{2}}(\beta_n-\beta)\right)\right\|_\infty \in [r,\infty)\right) \\ \leq \mathbb{P}\left(\|\mathrm{Arginf}(Z)\|_\infty \cap [r,\infty) \neq \emptyset\right) = \mathbb{P}(\gamma \geq r) = 1 - G(r-) \,.$$

For all $r > q_\eta$ we have that $G(r-) > 1-\eta$, see [Wit85, Lemma 1.17 b)]. Finally,

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}((\tau,\alpha,\beta) \in I_n(r)) \\
&= \liminf_{n\to\infty} \mathbb{P}\left(\left\|n(\tau_n-\tau), n^{\frac{1}{2}}(\alpha_n-\alpha), n^{\frac{1}{2}}(\beta_n-\beta)\right\|_\infty < r\right) \\
&= 1 - \limsup_{n\to\infty} \mathbb{P}\left(\left\|n(\tau_n-\tau), n^{\frac{1}{2}}(\alpha_n-\alpha), n^{\frac{1}{2}}(\beta_n-\beta)\right\|_\infty \in [r,\infty)\right) \\
&\geq G(r-) \\
&> 1-\eta \,.
\end{aligned}$$

□

We complete the chapter and apply the results to Example 6.3 from Section 6.1.

Example 6.20 (Sequel of Example 6.3) We use the results from 6.3 and derive

$$\mathbb{P}_{g(Y)|X}(\tau-,\cdot) \stackrel{6.1}{=} \mathbb{P}_{Y|X}\left(\tau-, g^{-1}(\cdot)\right) = \mathbb{P}\circ(\epsilon+\alpha)^{-1}\circ g^{-1} = \mathbb{P}\circ(g(\epsilon+\alpha))^{-1} \,,$$

and, analogously,

$$\mathbb{P}_{g(Y)|X}(\tau+,\cdot) = \mathbb{P}\circ(g(\epsilon+\beta))^{-1} \,.$$

Thus if $\xi_i' \sim \mathbb{P}_{Y-\frac{\alpha+\beta}{2}|X}(\tau+,\cdot)$ and $\xi_i \sim \mathbb{P}_{\frac{\alpha+\beta}{2}-Y|X}(\tau-,\cdot)$ then

$$\xi_i' \stackrel{\mathscr{L}}{=} \epsilon_i + \frac{\beta-\alpha}{2} \text{ and } \xi_i \stackrel{\mathscr{L}}{=} \frac{\beta-\alpha}{2} - \epsilon_i \,.$$

Furthermore,

$$\mathbb{E}\left(\mathbb{1}_{X\leq\tau}(Y-\alpha)^2\right) \stackrel{(2.10)}{=} \int_{(-\infty,\tau]} \int (y-m(x))^2 \mathbb{P}_{Y|X}(x,\mathrm{d}y)F(\mathrm{d}x) = \mathbb{E}(Y-\alpha)^2 F(\tau)$$

and, analogously, $\mathbb{E}(\mathbb{1}_{X\le\tau}(Y-\beta)^2) = \mathbb{E}(Y-\beta)^2\bar{F}(\tau)$. Hence, by Theorem 6.17

$$\limsup_{n\to\infty} \mathbb{P}\Big(\big(n(\tau_n-\tau), n^{\frac{1}{2}}(\alpha_n-\alpha), n^{\frac{1}{2}}(\beta_n-\beta)\big) \in C\Big) \le \mathbb{P}(\text{Arginf}(Z)\cap C \neq \emptyset)$$

for all closed $C \subseteq \mathbb{R}^3$, with

$$\text{Arginf}(Z) = \operatorname*{Arginf}_{t\in\mathbb{R}}\big(2(\beta-\alpha)Z^{(1)}(t)\big) \times \left\{\sqrt{\frac{\mathbb{E}\,(Y-\alpha)^2}{F(\tau)}}W\right\} \times \left\{\sqrt{\frac{\mathbb{E}\,(Y-\beta)^2}{\bar{F}(\tau)}}W'\right\},$$

where W and W' are two independent standard normal distributed random variables independent of the process

$$2\,(\beta-\alpha)\,Z^{(1)}(t) = \begin{cases} \sum\limits_{i=1}^{N_1(t)} 2\,(\beta-\alpha)\,\epsilon_i + (\alpha-\beta)^2 & ,t\ge 0 \\ \sum\limits_{i=1}^{N_2(-t)} 2\,(\alpha-\beta)\,\epsilon_i + (\alpha-\beta)^2 & ,t<0 \end{cases}$$

with the corresponding jump intensities $F'_-(\tau)$ and $F'_+(\tau+)$. This result is consistent with the limit process stated in [Kos08, p. 276], where we shall indicate that the author defined the estimator as a maximizer what causes a change of sign. [1]

[1]The intensity parameter there was stated as the density of ϵ, a typographical error.

7 Continuous Case

In this chapter the same methods as in chapter 6 are utilized to prove distributional convergence for the estimator in the case of regression functions which are continuous at the split point. In Section 7.1 we formulate additionally required conditions. In Section 7.2 the rescaled process is introduced and its convergence to the limit process, which was already identified in [BM07, Theorem 2.1], is proven. By means of the stochastic boundedness, which is proven in Section 7.3, we are able to prove the main theorem of this chapter in Section 7.4.

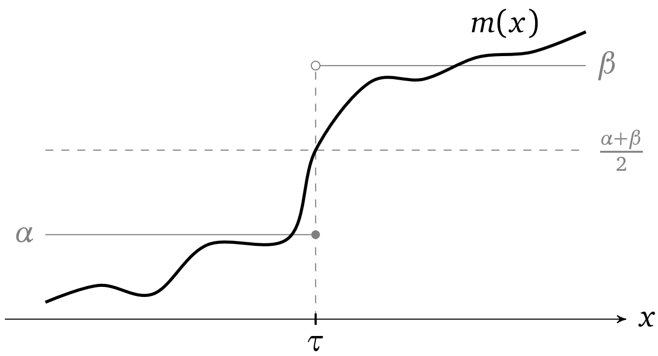
7.1 m is continuous at τ

Condition (A2) will now be further restricted.

(A2a) There exists an open ε-neighborhood $U_\varepsilon(\tau)$ of τ such that F is twice continuously differentiable in $U_\varepsilon(\tau)$ with $F'(\tau) > 0$.

(C1) The regression function m is continuously differentiable in an open ε-neighborhood $U_\varepsilon(\tau)$ of τ, where

$$\frac{m'(\tau)}{\beta - \alpha} > \frac{1}{4} \frac{F'(\tau)}{F(\tau) - F(\tau)^2} \,. \tag{7.1}$$



(C2) There exists an open ε-neighborhood $U_\varepsilon(\tau)$ of τ and $p > 2$ such that

$$\sup_{x \in U_\varepsilon(\tau)} \mathbb{E}\left(|Y|^p | X = x\right) < \infty \,.$$

Note that condition (7.1) is sufficient but not necessary to ensure that (τ, α, β) minimizes the criterion function S, cf. Lemma 2.1(vii).

7.2 Weak convergence of the rescaled process

We define the rescaled process for the continuous case

$$\tilde{Z}_n(t, a, b) := n^{\frac{2}{3}} \left\{ S_n\left(\tau + n^{-\frac{1}{3}} t, \alpha + n^{-\frac{1}{3}} a, \beta + n^{-\frac{1}{3}} b\right) - S_n(\tau, \alpha, \beta) \right\} \,. \tag{7.2}$$

Using the decompostion in (2.13) leads to

$$\tilde{Z}_n(t, a, b) = 2(\beta - \alpha)\tilde{Z}_n^{(1)}(t) + \tilde{Z}_n^{(2)}(t, a, b) \,,$$

where we set

$$\begin{aligned}
\tilde{Z}_n^{(1)}(t) &= n^{-\frac{1}{3}} \sum_{i=1}^n \left(\mathbb{1}_{X_i \le \tau + n^{-\frac{1}{3}} t} - \mathbb{1}_{X_i \le \tau}\right)\left(Y_i - \frac{\alpha + \beta}{2}\right) \\
\tilde{Z}_n^{(2)}(t, a, b) &= 2an^{-\frac{2}{3}} \sum_{i=1}^n \mathbb{1}_{X_i \le \tau}(\alpha - Y_i) + 2bn^{-\frac{2}{3}} \sum_{i=1}^n \mathbb{1}_{X_i > \tau}(\beta - Y_i) \\
&\quad + 2(b - a)n^{-\frac{2}{3}} \sum_{i=1}^n \left(\mathbb{1}_{X_i \le \tau + n^{-\frac{1}{3}} t} - \mathbb{1}_{X_i \le \tau}\right)\left(Y_i - \frac{\alpha + \beta}{2}\right) \\
&\quad + (a + b)(\alpha - \beta)n^{-\frac{2}{3}} \sum_{i=1}^n \left(\mathbb{1}_{X_i \le \tau + n^{-\frac{1}{3}} t} - \mathbb{1}_{X_i \le \tau}\right) \\
&\quad + a^2 n^{-1} \sum_{i=1}^n \mathbb{1}_{X_i \le \tau + n^{-\frac{1}{3}} t} + b^2 n^{-1} \sum_{i=1}^n \mathbb{1}_{X_i > \tau + n^{-\frac{1}{3}} t} \,.
\end{aligned} \tag{7.3}$$

Further notations are introduced

$$\begin{aligned}
\tilde{\Delta}_n(t, a, b) &:= \mathbb{E}\tilde{Z}_n(t, a, b) \\
&= 2(\beta - \alpha)\mathbb{E}\tilde{Z}_n^{(1)}(t) + \mathbb{E}\tilde{Z}_n^{(2)}(t, a, b) \\
&=: 2(\beta - \alpha)\tilde{\Delta}_n^{(1)}(t) + \tilde{\Delta}_n^{(2)}(t, a, b) \,,
\end{aligned}$$

where we compute

$$\tilde{\Delta}_n^{(1)}(t) = n^{\frac{2}{3}}\left(\left(H(\tau+n^{-\frac{1}{3}}t)-H(\tau)\right)-\frac{\alpha+\beta}{2}\left(F(\tau+n^{-\frac{1}{3}}t)-F(\tau)\right)\right). \tag{7.4}$$

Under assumption (A1) we can use the normal equations in Lemma 2.1(iii) to see that

$$\begin{aligned}\tilde{\Delta}_n^{(2)}(t,a,b) = 2(b-a)n^{\frac{1}{3}}&\left(\left(H(\tau+n^{-\frac{1}{3}}t)-H(\tau)\right)-\frac{\alpha+\beta}{2}\left(F(\tau+n^{-\frac{1}{3}}t)-F(\tau)\right)\right)\\ &+(a+b)(\alpha-\beta)\,n^{\frac{1}{3}}\left(F(\tau+n^{-\frac{1}{3}}t)-F(\tau)\right)\\ &+a^2F(\tau+n^{-\frac{1}{3}}t)+b^2\bar{F}(\tau+n^{-\frac{1}{3}}t)\,. \end{aligned}\tag{7.5}$$

Now we define the limiting process of $\tilde{Z}_n$.

Definition 7.1 Let $(\tilde{Z}^{(1)}(t))_{t\in\mathbb{R}} \in D(\mathbb{R})$ be a two-sided Brownian motion with quadratic drift of the form

$$\tilde{Z}^{(1)}(t) = V(\tau)F'(\tau)B(t)+\frac{1}{2}m'(\tau)F'(\tau)t^2\,,$$

where $(B(t))_{t\in\mathbb{R}}$ is a two-sided standard Brownian motion in $D(\mathbb{R})$. Furthermore, let $\tilde{Z}^{(2)}$ be a deterministic function in $D(\mathbb{R}^3)$ with

$$\tilde{Z}^{(2)}(t,a,b) = (\alpha-\beta)F'(\tau)(a+b)t+a^2F(\tau)+b^2\bar{F}(\tau)\,.$$

Now introduce the stochastic process $\tilde{Z} \in D(\mathbb{R}^3)$ with

$$\tilde{Z}(t,a,b) = 2(\beta-\alpha)\tilde{Z}^{(1)}(t)+\tilde{Z}^{(2)}(t,a,b)\,.$$

The remaining part of this section is dedicated to the proof of the main statement below. It will be proven at the end of this section by virtue of the Lemmas given now.

Theorem 7.2 If (A1)-(A3) and (C1)-(C2) hold, then

$$\tilde{Z}_n \xrightarrow{\mathscr{L}} \tilde{Z} \quad \text{in } D(\mathbb{R}^3)\,,$$

as $n \to \infty$.

In preparation we prove the distributional convergence of $2(\beta-\alpha)\tilde{Z}_n^{(1)}$ to the Brownian Motion with quadratic drift.

Lemma 7.3 If (A1)-(A3) and (C1)-(C2) hold, then

$$2(\beta-\alpha)\tilde{Z}_n^{(1)} \xrightarrow{\mathscr{L}} 2(\beta-\alpha)\tilde{Z}^{(1)} \quad \text{in } D(\mathbb{R})\,, \tag{7.6}$$

as $n \to \infty$.

Proof. Decompose $(Y_i - {}^{\alpha+\beta}/_2) = (Y_i - m(X_i)) + (m(X_i) - {}^{\alpha+\beta}/_2)$ and write $\tilde{Z}_n^{(1)} = \Gamma_n + \bar{\Gamma}_n$, where

$$\Gamma_n(t) = n^{-\frac{1}{3}} \sum_{i=1}^{n} \left(\mathbb{1}_{X_i \leq \tau + n^{-\frac{1}{3}} t} - \mathbb{1}_{X_i \leq \tau} \right) (Y_i - m(X_i))$$

and

$$\bar{\Gamma}_n(t) = n^{-\frac{1}{3}} \sum_{i=1}^{n} \left(\mathbb{1}_{X_i \leq \tau + n^{-\frac{1}{3}} t} - \mathbb{1}_{X_i \leq \tau} \right) \left(m(X_i) - \frac{\alpha+\beta}{2} \right).$$

Set $d > 0$. From [Fer09, Section 8.1] we know that $\Gamma_n \xrightarrow{\mathscr{L}} \Gamma$, where

$$\Gamma(t) = \begin{cases} V(\tau)F'(\tau)B_1(t) & , t \geq 0 \\ V(\tau)F'(\tau)B_2(-t) & , t < 0 \end{cases} \quad \text{in } D[-d,d]\,,$$

where B_1 and B_2 are both Brownian Motions and B_1 and B_2 are independent. Next, derive

$$\mathbb{E}\bar{\Gamma}_n(t) = \begin{cases} n^{\frac{2}{3}} \displaystyle\int_{(\tau, \tau + n^{-\frac{1}{3}} t]} m(x) - m(\tau)\, F(\mathrm{d}x) & , t \geq 0 \\ -n^{\frac{2}{3}} \displaystyle\int_{(\tau + n^{-\frac{1}{3}} t, \tau]} m(x) - m(\tau)\, F(\mathrm{d}x) & , t < 0 \end{cases}.$$

Since m is continuously differentiable in an open neighborhood $U_\varepsilon(\tau)$ of τ, $m(\tau) = {}^1/_2(\alpha+\beta)$ (cf. Lemma 2.1(vi)), and for each $x \in [\tau, \tau+\varepsilon]$ there exists an $\xi \in [\tau, x]$ such that $m(x) - m(\tau) = m'(\xi)(x-\tau)$. A successive application of the Integration by parts theorem ([HS75, Theorem 21.67]), the substitution $u = \tau + n^{-1/3}t$, the L'Hospital's rule, and the fundamental theorem of calculus then yields

$$
\begin{aligned}
&\lim_{n\to\infty} n^{\frac{2}{3}} \int\limits_{(\tau,\tau+n^{-\frac{1}{3}}t]} m(x)-m(\tau)\,F(\mathrm{d}x)\\
&\le \lim_{n\to\infty} \sup_{\xi\in(\tau,\tau+n^{-\frac{1}{3}}t]} m'(\xi)\,n^{\frac{2}{3}} \int\limits_{(\tau,\tau+n^{-\frac{1}{3}}t]} x-\tau\,F(\mathrm{d}x)\\
&= \lim_{n\to\infty} \sup_{\xi\in(\tau,\tau+n^{-\frac{1}{3}}t]} m'(\xi)\,n^{\frac{2}{3}} \left(F(\tau+n^{-\frac{1}{3}}t)n^{-\frac{1}{3}}t - \int\limits_{(\tau,\tau+n^{-\frac{1}{3}}t]} F(x)\,\mathrm{d}x \right)\\
&= t^2 \lim_{u\downarrow\tau} \sup_{\xi\in(\tau,u]} m'(\xi) \frac{F(u)(u-\tau)-\int_{(\tau,u]} F(x)\,\mathrm{d}x}{(u-\tau)^2}\\
&= \frac{1}{2} t^2 m'(\tau)F'(\tau)\,.
\end{aligned}
$$

Analogously, we find $\frac{1}{2}t^2 m'(\tau)F'(\tau)$ as a lower bound by using $\inf_{\xi\in[\tau,\tau+tn^{-\frac{1}{3}}]} m'(\xi)$ instead of $\sup_{\xi\in[\tau,\tau+tn^{-\frac{1}{3}}]} m'(\xi)$. In the same way we obtain an estimate for $\mathbb{E}\bar{\Gamma}_n(t)$ for all $t<0$ and, therefore,

$$
\lim_{n\to\infty} \mathbb{E}\bar{\Gamma}_n(t) = \frac{1}{2}t^2 m'(\tau)F'(\tau) =: \bar{\Gamma}(t)
$$

follows from the continuous differentiability of m. To evaluate the variance we proceed in much the same way and obtain for all $t\in[0,d]$

$$
\begin{aligned}
&\mathrm{Var}\bar{\Gamma}_n(t)\\
&\le n^{\frac{1}{3}} \int\limits_{(\tau,\tau+n^{-\frac{1}{3}}t]} (m(x)-m(\tau))^2\,F(\mathrm{d}x)\\
&\le n^{\frac{1}{3}} \sup_{\xi\in[\tau,\tau+n^{-\frac{1}{3}}t]} \left(m'(\xi)\right)^2 \int\limits_{(\tau,\tau+n^{-\frac{1}{3}}t]} (x-\tau)^2\,F(\mathrm{d}x)\\
&= n^{\frac{1}{3}} \sup_{\xi\in[\tau,\tau+n^{-\frac{1}{3}}t]} \left(m'(\xi)\right)^2 \int\limits_{(\tau,\tau+n^{-\frac{1}{3}}t]} \Big(F(\tau+n^{-\frac{1}{3}}t)-F(x)\Big)2(x-\tau)\,\mathrm{d}x\\
&\le n^{-\frac{1}{3}} \sup_{\xi\in[\tau,\tau+n^{-\frac{1}{3}}t]} \left(m'(\xi)\right)^2 \Big(F(\tau+n^{-\frac{1}{3}}t)-F(\tau)\Big)t^2\,,
\end{aligned}
$$

where we used the Integration by parts theorem ([HS75, Theorem 21.67]) in the next to last step. Using similar arguments when $t \in [-d,0]$, we obtain $\lim_{n\to\infty} \mathrm{Var}\bar{\Gamma}_n(t) = 0$ for all $t \in [-d,d]$. By Chebyshev's inequality it follows that $\bar{\Gamma}_n(t) \xrightarrow{\mathbb{P}} \bar{\Gamma}(t)$ in $[-d,d]$, and by Slutsky's theorem ([GS77, Proposition 8.6.4]) then $(\Gamma_n, \bar{\Gamma}_n) \xrightarrow{\mathscr{L}} (\Gamma, \bar{\Gamma})$ in $D[-d,d]\times D[-d,d]$. Now we use the fact that addition $(f,g) \mapsto f+g$ is s_1-continuous in $D[-d,d]$ if one summand is continuous itself (cf. [Fer09, Lemma 2.3 (3)]). Together with the CMT we infer $2(\beta-\alpha)\tilde{Z}_n^{(1)} \xrightarrow{\mathscr{L}} 2(\beta-\alpha)\tilde{Z}^{(1)}$ in $D[-d,d]$ for all $d>0$. Finally, Proposition 3.5 proofs the claim. □

Lemma 7.4 If (A1)-(A2) and (C1) hold, then for all $d>0$

$$\left\|\tilde{Z}_n^{(2)} - \tilde{\Delta}_n^{(2)}\right\|_{[-d,d]^3} \xrightarrow{\mathbb{P}} 0\,, \tag{7.7}$$

as $n \to \infty$.

Proof. Use (7.3) and compute

$$\begin{aligned}
&\sup_{(t,a,b)\in[-d,d]^3} \left|\tilde{Z}_n^{(2)}(t,a,b) - \tilde{\Delta}_n^{(2)}(t,a,b)\right| \\
&\leq 2d\left|n^{-\frac{2}{3}}\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau}(\alpha - Y_i)\right| + 2d\left|n^{-\frac{2}{3}}\sum_{i=1}^n \mathbb{1}_{X_i>\tau}(\beta - Y_i)\right| \\
&\qquad + 4d\sup_{t\in[-d,d]}\left|n^{-\frac{2}{3}}\sum_{i=1}^n\left(\left(\mathbb{1}_{X_i\leq\tau+n^{-\frac{1}{3}}t} - \mathbb{1}_{X_i\leq\tau}\right)\left(Y_i - \frac{\alpha+\beta}{2}\right)\right.\right. \\
&\qquad\qquad \left.\left. -\mathbb{E}\left[\left(\mathbb{1}_{X_i\leq\tau+n^{-\frac{1}{3}}t} - \mathbb{1}_{X_i\leq\tau}\right)\left(Y_i - \frac{\alpha+\beta}{2}\right)\right]\right)\right| \\
&+2(|\alpha|+|\beta|)d\sup_{t\in[-d,d]}\left|n^{-\frac{2}{3}}\sum_{i=1}^n\left(\left(\mathbb{1}_{X_i\leq\tau+n^{-\frac{1}{3}}t} - \mathbb{1}_{X_i\leq\tau}\right) - \mathbb{E}\left(\mathbb{1}_{X_i\leq\tau+n^{-\frac{1}{3}}t} - \mathbb{1}_{X_i\leq\tau}\right)\right)\right| \\
&\qquad + 2d^2\sup_{t\in[-d,d]}\left|n^{-1}\sum_{i=1}^n\left(\mathbb{1}_{X_i\leq\tau+n^{-\frac{1}{3}}t} - \mathbb{E}\,\mathbb{1}_{X_i\leq\tau+n^{-\frac{1}{3}}t}\right)\right|\,.
\end{aligned} \tag{7.8}$$

The sums in the first term on the right hand side of Inequality (7.8) have i. i. d. zero-mean and square-integrable summands as was mentioned in Lemma 2.1(iii). By the CLT one has $n^{-1/2}\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau}(\alpha - Y_i) \xrightarrow{\mathscr{L}} \mathcal{N}(0, T(\tau))$ and, hence, $2d\left|n^{-2/3}\sum_{i=1}^n \mathbb{1}_{X_i\leq\tau}(\alpha - Y_i)\right| \xrightarrow{\mathbb{P}} 0$, as $n\to\infty$. With similar arguments $2d\left|n^{-2/3}\sum_{i=1}^n \mathbb{1}_{X_i>\tau}(\beta - Y_i)\right| \xrightarrow{\mathbb{P}} 0$, as $n\to\infty$. Consider the second line on the right hand side of (7.8) and recall the notion of the process W_n in

(4.5). Then we have for all $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[-d,d]}\left|n^{-\frac{2}{3}}\sum_{i=1}^{n}\left(\left(\mathbb{1}_{X_i\le\tau+n^{-\frac{1}{3}}t}-\mathbb{1}_{X_i\le\tau}\right)\left(Y_i-\frac{\alpha+\beta}{2}\right)\right.\right.\right.\\
&\qquad\qquad\left.\left.\left.-\mathbb{E}\left[\left(\mathbb{1}_{X_i\le\tau+n^{-\frac{1}{3}}t}-\mathbb{1}_{X_i\le\tau}\right)\left(Y_i-\frac{\alpha+\beta}{2}\right)\right]\right)\right|\ge\varepsilon\right)\\
&=\mathbb{P}\left(\bigcup_{|t|\le d}\left\{n^{\frac{1}{3}}\left|W_n(\tau+n^{-\frac{1}{3}}t)-W_n(\tau)\right|\ge\varepsilon\right\}\right)\\
&\longrightarrow 0\,,
\end{aligned}$$

as $n\to\infty$, by Lemma 4.18, with $\nu = 1/3$, $\lambda = 1$ and $\gamma = 0$. Consequently, the second term on the right hand side of Equation (7.8) converges in probability towards zero. Moreover, it follows that the third term converges in probability towards zero, since the process in vertical bars is one of the processes occurring in the decomposition of W_n in (4.22). To see the convergence of the last term note that

$$\sup_{t\in[-d,d]}\left|n^{-1}\sum_{i=1}^{n}\left(\mathbb{1}_{X_i\le\tau+n^{-\frac{1}{3}}t}-\mathbb{E}\,\mathbb{1}_{X_i\le\tau+n^{-\frac{1}{3}}t}\right)\right|\le\sup_{t\in\mathbb{R}}|F_n(t)-F(t)|$$

which converges towards zero almost surely by the Glivenko-Cantelli Theorem (cf. [SW86, Theorem 3.3]). □

Lemma 7.5 If (A1)-(A2) and (C1) hold, then for all $d > 0$

$$\left\|\tilde{\Delta}_n^{(2)}-\tilde{Z}^{(2)}\right\|_{[-d,d]^3}\longrightarrow 0\,, \tag{7.9}$$

as $n\to\infty$.

Proof. We state the proof for the case when $\alpha < \beta$. The case $\beta < \alpha$ may be handled similar. Set $d > 0$ and choose $\varepsilon_1 > 0$ satisfying the conditions (A2) and (C1). By (C1) we could find $0 < \varepsilon \le \varepsilon_1$ such that m is monotonically increasing on $U_\varepsilon(\tau)$. For all $|t| \le d$ and $n > (|d|/\varepsilon)^3$ then $\tau + n^{-1/3}t \in U_\varepsilon(\tau)$. Write

$$\begin{aligned}
h_n(t) :=& n^{\frac{1}{3}}\left(\left(H(\tau+n^{-\frac{1}{3}}t)-H(\tau)\right)-\frac{\alpha+\beta}{2}\left(F(\tau+n^{-\frac{1}{3}}t)-F(\tau)\right)\right)\\
=&\begin{cases} n^{\frac{1}{3}}\int_{(\tau,\tau+n^{-\frac{1}{3}}t]} m(x)-\frac{\alpha+\beta}{2}F(\mathrm{d}x) & ,t\ge 0\\ n^{\frac{1}{3}}\int_{(\tau+n^{-\frac{1}{3}}t,\tau]} \frac{\alpha+\beta}{2}-m(x)F(\mathrm{d}x) & ,t<0\end{cases}
\end{aligned}$$

and observe that for all $n > (|d|/\varepsilon)^3$ and $0 \le t \le d$

$$\begin{aligned}
0 &\le h_n(t) \\
&= n^{\frac{1}{3}} \int_{(\tau,\tau+n^{-\frac{1}{3}}t]} m(x) - m(\tau) F(\,\mathrm{d}x) \\
&\le n^{\frac{1}{3}} \Big(m(\tau + n^{-\frac{1}{3}} t) - m(\tau) \Big) \Big(F(\tau + n^{-\frac{1}{3}} t) - F(\tau) \Big) \\
&\le \Big(m(\tau + n^{-\frac{1}{3}} d) - m(\tau) \Big) \frac{F(\tau + n^{-\frac{1}{3}} d) - F(\tau)}{n^{-\frac{1}{3}} d} d \,.
\end{aligned}$$

If $-d \le t < 0$, then for all $n > (|d|/\varepsilon)^3$

$$0 \le h_n(t) \le \Big(m(\tau) - m(\tau - n^{-\frac{1}{3}} d) \Big) \frac{F(\tau) - F(\tau - n^{-\frac{1}{3}} d)}{n^{-\frac{1}{3}} d} d \,.$$

Using (A2) and the continuity of m at τ, we conclude

$$\sup_{t \in [-d,d]} |h_n(t)| \longrightarrow 0 \,, \tag{7.10}$$

as $n \to \infty$. Denote

$$g_n(t) := n^{\frac{1}{3}} \Big(\Big(F(\tau + n^{-\frac{1}{3}} t) - F(\tau) \Big) - F'(\tau) t \Big)$$

and use similar arguments to see that

$$\sup_{t \in [-d,d]} |g_n(t)| \longrightarrow 0 \,, \tag{7.11}$$

as $n \to \infty$. Finally, use (7.5), (7.10) and (7.10) to see that

$$\begin{aligned}
&\sup_{(t,a,b) \in [-d,d]^3} \Big| \tilde{\Delta}_n^{(2)}(t,a,b) - \tilde{Z}^{(2)}(t,a,b) \Big| \\
&= \sup_{(t,a,b) \in [-d,d]^3} \Big| 2(b-a) h_n(t) + (a+b)(\alpha - \beta)\, g_n(t) \\
&\qquad\qquad + \big(a^2 - b^2\big) \Big[F(\tau + n^{-\frac{1}{3}} t) - F(\tau) \Big] \Big| \\
&\le 4d \sup_{t \in [-d,d]} |h_n(t)| + 2d\, |\alpha - \beta| \sup_{t \in [-d,d]} |g_n(t)| + 2d^2 \Big(F(\tau + n^{-\frac{1}{3}} d) - F(\tau - n^{-\frac{1}{3}} d) \Big) \\
&\longrightarrow 0 \,,
\end{aligned}$$

as $n \to \infty$. □

Proof of Theorem 7.2. Set $d > 0$. Use Lemma 3.4 and the map Φ introduced therein to define a process $(\bar{Z}_n^{(1)})_{n\in\mathbb{N}} \in D([-d,d])^3$ with $\bar{Z}_n^{(1)} := \Phi(\tilde{Z}_n^{(1)})$. Then Φ is continuous and by Lemma 7.3 and the CMT then $\bar{Z}_n^{(1)} \xrightarrow{\mathcal{L}} \bar{Z}^{(1)}$ in $\Phi(D[-d,d])$, where $\bar{Z}^{(1)}(t,a,b) = \tilde{Z}^{(1)}(t)$. By [Kal97, Lemma 3.26] we also have that $\bar{Z}_n^{(1)} \xrightarrow{\mathcal{L}} \bar{Z}^{(1)}$ in $D([-d,d]^3)$. Slutsky's theorem ([GS77, Proposition 8.6.4]) and the CMT then imply

$$2(\beta-\alpha)\tilde{Z}_n^{(1)}(t) + \tilde{Z}^{(2)}(t,a,b) \xrightarrow{\mathcal{L}} 2(\beta-\alpha)\tilde{Z}^{(1)}(t) + \tilde{Z}^{(2)}(t,a,b) \qquad \text{in } D([-d,d])^3 \,.$$

Note that Lemmas 7.4 and 7.5 imply

$$\begin{aligned}
&\sup_{(t,a,b)\in[-d,d]^3} \left|\tilde{Z}_n(t,a,b) - \left(2(\beta-\alpha)\tilde{Z}_n^{(1)}(t) + \tilde{Z}^{(2)}(t,a,b)\right)\right| \\
&= \sup_{(t,a,b)\in[-d,d]^3} \left|\tilde{Z}_n^{(2)}(t,a,b) - \tilde{Z}^{(2)}(t,a,b)\right| \\
&\le \sup_{(t,a,b)\in[-d,d]^3} \left|\tilde{Z}_n^{(2)}(t,a,b) - \tilde{\Delta}_n^{(2)}(t,a,b)\right| + \sup_{(t,a,b)\in[-d,d]^3} \left|\tilde{\Delta}_n^{(2)}(t,a,b) - \tilde{Z}^{(2)}(t,a,b)\right| \\
&\xrightarrow{\mathbb{P}} 0 \,,
\end{aligned}$$

as $n \to \infty$. Finally, by Slutsky's theorem and Proposition 3.5, $\tilde{Z}_n \xrightarrow{\mathcal{L}} \tilde{Z}$ in $D(\mathbb{R}^3)$. □

7.3 Stochastic boundedness of the estimator

Theorem 7.6 Suppose that assumptions **(A1)**, **(A2a)**, **(A3)** and **(C1)** hold and $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$, as $n \to \infty$, then

$$\lim_{d\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\left\| n^{\frac{1}{3}} \begin{pmatrix} \tau_n - \tau \\ \alpha_n - \alpha \\ \beta_n - \beta \end{pmatrix} \right\| \ge d \right) = 0 \,.$$

Lemma 7.7 Under the assumptions of **(A1)**, **(A2a)**, **(A3)** and **(C1)** there exists some $\delta > 0$ and a constant $D > 0$ such that

$$\tilde{\Delta}_n(t,a,b) \ge D \, \|(t,a,b)\|_\infty^2$$

for all (t,a,b) with $\|n^{-1/3}(t,a,b)\|_\infty \le \delta$.

Proof. Set $y=(t,a,b)$ and $\theta=(\tau,\alpha,\beta)$, then $\tilde{\Delta}_n(y)=n^{2/3}\{S(\theta+n^{-\frac{1}{3}}y)-S(\theta)\}$. Since by condition (A1) S obtains its minimum at θ, $\nabla S(\theta)=0$. By Taylor's theorem for multivariate functions, see [Heu91, Proposition 168.2], we can write

$$\tilde{\Delta}_n(y)=n^{\frac{2}{3}}\left\{\frac{1}{2}n^{-\frac{2}{3}}y^\top \mathrm{H}_S(\theta)y+\rho(n^{-\frac{1}{3}}y)\left\|n^{-\frac{1}{3}}y\right\|^2\right\}=\frac{1}{2}y^\top \mathrm{H}_S(\theta)y+\rho(n^{-\frac{1}{3}}y)\left\|y\right\|^2,$$

where $\rho(n^{-1/3}y)\to 0$, as $n\to\infty$, and H_S denotes the Hessian matrix of S. In order to obtain $\mathrm{H}_S(\theta)$, we compute

$$\nabla S(y)=\left((a^2-b^2)F'(t)+2(b-a)m(t)F'(t),2aF(t)-2H(t),2b\bar{F}(t)-2\bar{H}(t)\right)^\top,$$

where we used, that $H'(t)=m(t)F'(t)$ (cf. Equation (2.9)), and

$$\mathrm{H}_S(y)=$$
$$\begin{pmatrix} (b-a)[F''(t)(2m(t)-(a+b))+2m'(t)F'(t)] & 2F'(t)(a-m(t)) & 2F'(t)(m(t)-b) \\ 2F'(t)(a-m(t)) & 2F(t) & 0 \\ 2F'(t)(m(t)-b) & 0 & 2\bar{F}(t) \end{pmatrix}.$$

Since $2m(\tau)=\alpha+\beta$ by Lemma 2.1(vi), we get

$$\mathrm{H}_S(\theta)=\begin{pmatrix} 2(\beta-\alpha)m'(\tau)F'(\tau) & (\alpha-\beta)F'(\tau) & (\alpha-\beta)F'(\tau) \\ (\alpha-\beta)F'(\tau) & 2F(\tau) & 0 \\ (\alpha-\beta)F'(\tau) & 0 & 2\bar{F}(\tau) \end{pmatrix}$$

and its determinant

$$\det \mathrm{H}_S(\theta)=8(\beta-\alpha)m'(\tau)F'(\tau)F(\tau)\bar{F}(\tau)-2(\alpha-\beta)^2F'^2(\tau)\,.$$

Observe that by Lemma 2.1(i) and by condition (7.1) in assumption (C1), $\det \mathrm{H}_S(\tau,\alpha,\beta)$ and all other leading principal minors are strictly positive. Hence, $\mathrm{H}_S(\theta)$ is positive definite (see [Fis02, p. 327]) and by [Heu91, Lemma 173.2] there exists $\lambda>0$ such that $y^\top \mathrm{H}_S(\theta)y\geq\lambda\,\|y\|^2$ and, therefore,

$$\tilde{\Delta}_n(y)\geq\left(\frac{1}{2}\lambda+\rho(n^{-\frac{1}{3}}y)\right)\|y\|^2\,.$$

Now choose $\delta=\delta(\lambda)>0$ sufficiently small such that $|\rho(n^{-1/3}y)|\leq 1/4\,\lambda$ for all $n^{-1/3}\|y\|_\infty\leq$

δ. Then $\tilde{\Delta}_n(y) \geq {}^{\lambda}/_{4}\,\|y\|^2$ and the claim follows by the equivalence of norms in $\mathbb{R}^3$. $\square$

Proof of Theorem 7.6. Write $\theta_n := (\tau_n, \alpha_n, \beta_n)$, $\theta := (\tau, \alpha, \beta)$, and $y = (t, a, b)$, and let $D > 0$ be a positive generic constant (cf. Remark 4.1). Similar to the proof of Theorem 6.15, fix $d > 0$ and choose $\delta > 0$ being admissible in the sense of Lemma 7.7. Then

$$\mathbb{P}\left(\left\| n^{\frac{1}{3}} \begin{pmatrix} \tau_n - \tau \\ \alpha_n - \alpha \\ \beta_n - \beta \end{pmatrix} \right\| \geq d \right) \leq \mathbb{P}\left(n^{\frac{1}{3}}\,\|(\theta_n - \theta)\| \geq d\,, \|(\theta_n - \theta)\| \leq \delta\right) + \mathbb{P}\left(\|(\theta_n - \theta)\| > \delta\right). \tag{7.12}$$

By $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$ we already know that $\lim_{n\to\infty} \mathbb{P}(\|(\theta_n - \theta)\| > \delta) = 0$ for all $d > 0$. To handle the first term on the right hand side of (7.12) set

$$J := \left\{ y = (t, a, b);\ d \leq \|y\|_\infty \leq n^{\frac{1}{3}}\delta \right\}$$

and use Lemma 7.7 to deduce

$$\begin{aligned}
&\mathbb{P}\left(n^{\frac{1}{3}}\,\|(\theta_n - \theta)\| \geq d\,, \|(\theta_n - \theta)\| \leq \delta\right) \\
&\leq \mathbb{P}\left(\bigcup_{y\in J} \left\{ S_n\left(\theta + n^{-\frac{1}{3}} y\right) \leq S_n(\theta) \right\}\right) \\
&= \mathbb{P}\left(\bigcup_{y\in J} \{\tilde{Z}_n(y) \leq 0\}\right) \\
&= \mathbb{P}\left(\bigcup_{y\in J} \{\tilde{\Delta}_n(y) - \tilde{Z}_n(y) \geq \tilde{\Delta}_n(y)\}\right) \\
&\leq \mathbb{P}\left(\bigcup_{y\in J} \left\{\left|\tilde{\Delta}_n(y) - \tilde{Z}_n(y)\right| \geq D\,\|y\|^2\right\}\right) \\
&\leq \mathbb{P}\left(\bigcup_{(t,a,b)\in J} \left\{\left|2(\alpha - \beta)\left(\tilde{Z}_n^{(1)}(t) - \tilde{\Delta}_n^{(1)}(t)\right)\right| \geq D\,\|(t, a, b)\|^2\right\}\right) \\
&\qquad + \mathbb{P}\left(\bigcup_{(t,a,b)\in J} \left\{\left|\tilde{Z}_n^{(2)}(t, a, b) - \tilde{\Delta}_n^{(2)}(t, a, b)\right| \geq D\,\|(t, a, b)\|^2\right\}\right).
\end{aligned} \tag{7.13}$$

We adopt the splitting of J into $J_1 := \{y \in J; |t| \geq d\}$ and $J_2 := \{y \in J\,; |t| < d\}$ from the proof of Theorem 6.15. By recalling the notion of W_n in (4.5) we obtain the estimate

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|2(\alpha-\beta)\left(\tilde{Z}_n^{(1)}(t)-\tilde{\Delta}_n^{(1)}(t)\right)\right|\geq D\,\|(t,a,b)\|^2\right\}\right)\\
&\leq \mathbb{P}\left(\bigcup_{d\leq|t|\leq n^{\frac{1}{3}}\delta}\left\{n^{\frac{2}{3}}\left|W_n(\tau+n^{-\frac{1}{3}}t)-W_n(\tau)\right|\geq D\,|t|^2\right\}\right)\\
&\qquad\qquad\qquad + \mathbb{P}\left(\bigcup_{|t|<d}\left\{n^{\frac{2}{3}}\left|W_n(\tau+n^{-\frac{1}{3}}t)-W_n(\tau)\right|\geq Dd^2\right\}\right)\,.
\end{aligned}$$

Applying Lemmas 4.17 and 4.18, while choosing $\nu = 1/3$ and $\lambda = {}^{\nu+1}/_{2\nu}$, yields

$$\lim_{d\to\infty}\lim_{n\to\infty}\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|2(\alpha-\beta)\left(\tilde{Z}_n^{(1)}(t)-\tilde{\Delta}_n^{(1)}(t)\right)\right|\geq D\,\|(t,a,b)\|^2\right\}\right)=0\,.$$

Using (7.3) to estimate the second term in (7.13) yields

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{\left|\tilde{Z}_n^{(2)}(t,a,b)-\tilde{\Delta}_n^{(2)}(t,a,b)\right|\geq D\,\|(t,a,b)\|^2\right\}\right)\\
&\leq \mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\Bigg\{2\,(|a|\vee|b|)^2\left|F_n(\tau+n^{-\frac{1}{3}}t)-F(\tau+n^{-\frac{1}{3}}t)\right|\\
&+(|a|\vee|b|)\,n^{-\frac{2}{3}}\left(\left|\sum_{i=1}^{n}\mathbb{1}_{\{X_i\leq\tau\}}\,(\alpha-Y_i)\right|+\left|\sum_{i=1}^{n}\mathbb{1}_{\{X_i>\tau\}}\,(\beta-Y_i)\right|\right)\geq D\,\|(t,a,b)\|^2\Bigg\}\Bigg)\\
&\quad+\mathbb{P}\left(\bigcup_{(t,a,b)\in J}\left\{4\,(|a|\vee|b|)\,n^{\frac{1}{3}}\left|W_n(\tau+n^{-\frac{1}{3}}t)-W_n(\tau)\right|\geq D\,\|(t,a,b)\|^2\right\}\right)\\
&\qquad+\mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J}\Bigg\{(|a|\vee|b|)\,(|\alpha|+|\beta|)\,n^{-\frac{2}{3}}\Bigg|\sum_{i=1}^{n}\Big(\left(\mathbb{1}_{X_i\leq\tau+n^{-\nu}t}-\mathbb{1}_{X_i\leq\tau}\right)\\
&\qquad\qquad-\mathbb{E}\left(\mathbb{1}_{X_i\leq\tau+n^{-\nu}t}-\mathbb{1}_{X_i\leq\tau}\right)\Big)\Bigg|\geq D\,\|(t,a,b)\|^2\Bigg\}\Bigg)\,. \qquad (7.14)
\end{aligned}$$

Note that by Lemma 2.1(iii) and the SLLN

$$
\begin{aligned}
n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le \tau\}} (\alpha - Y_i) &\xrightarrow{\mathscr{L}} \mathscr{N}(0, T(\tau)) \,, \\
n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \tau\}} (\beta - Y_i) &\xrightarrow{\mathscr{L}} \mathscr{N}(0, \bar{T}(\tau)) \,,
\end{aligned}
$$

as $n \to \infty$. The quantity $\sqrt{n} \sup_{t \in [\tau-\delta, \tau+\delta]} |F_n(t) - F(t)|$ converges in distribution as $n \to \infty$ as a consequence of Donsker's theorem (cf. [JS03, Corollary 3.11]) and the CMT. Therefore,

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\Bigg(\bigcup_{(t,a,b)\in J} \Big\{ 2\,(|a| \vee |b|)^2 \left| F_n(\tau + n^{-\frac{1}{3}} t) - F(\tau + n^{-\frac{1}{3}} t) \right| \\
&\qquad + (|a| \vee |b|)\, n^{-\frac{2}{3}} \left(\left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le \tau\}} (\alpha - Y_i) \right| + \left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \tau\}} (\beta - Y_i) \right| \right) \ge D \,\|(t,a,b)\|^2 \Big\} \Bigg) \\
&\le \limsup_{n\to\infty} \mathbb{P}\left(\sup_{t\in[\tau-\delta,\tau+\delta]} |F_n(t) - F(t)| \ge D \right) \\
&\qquad\qquad + \limsup_{n\to\infty} \mathbb{P}\left(n^{-\frac{2}{3}} \left(\left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le \tau\}} (\alpha - Y_i) \right| + \left| \sum_{i=1}^{n} \mathbb{1}_{\{X_i > \tau\}} (\beta - Y_i) \right| \right) \ge Dd \right) \\
&= 0
\end{aligned}
$$

for all $d > 0$. Consider the second probability in Inequality (7.14) and use the splitting $J = J_1 \cup J_2$

$$
\begin{aligned}
&\mathbb{P}\left(\bigcup_{(t,a,b)\in J} 4\,(|a| \vee |b|)\, n^{\frac{1}{3}} \left| W_n(\tau + n^{-\frac{1}{3}} t) - W_n(\tau) \right| \ge D \,\|(t,a,b)\|^2 \right) \\
&\le \mathbb{P}\left(\bigcup_{(t,a,b)\in J_1} n^{\frac{1}{3}} \left| W_n(\tau + n^{-\frac{1}{3}} t) - W_n(\tau) \right| \ge D \,\|(t,a,b)\| \right) \\
&\qquad\qquad + \mathbb{P}\left(\bigcup_{(t,a,b)\in J_2} n^{\frac{1}{3}} \left| W_n(\tau + n^{-\frac{1}{3}} t) - W_n(\tau) \right| \ge D \,\|(t,a,b)\| \right) \\
&\le \mathbb{P}\left(\bigcup_{d \le |t| \le n^{\frac{1}{3}}\delta} n^{\frac{1}{3}} \left| W_n(\tau + n^{-\frac{1}{3}} t) - W_n(\tau) \right| \ge D\,|t| \right) \\
&\qquad\qquad + \mathbb{P}\left(\bigcup_{|t|<d} n^{\frac{1}{3}} \left| W_n(\tau + n^{-\frac{1}{3}} t) - W_n(\tau) \right| \ge Dd \right) .
\end{aligned}
$$

Applying Lemmas 4.17 and 4.18, with $\nu = 1/3$, $\lambda = 1$ and $\gamma = 1$, leads to

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{(t,a,b)\in J} 4\,(|a|\vee|b|)\,n^{\frac{1}{3}} \left| W_n(\tau + n^{-\frac{1}{3}}t) - W_n(\tau) \right| \geq D\,\|(t,a,b)\|^2 \right) = 0\,,$$

for all $d > 0$. Since the process in the third probability on the right hand side of (7.14) is one of the processes occurring in the decomposition of W_n in (4.25) (the corresponding statements are contained in the proofs of Lemmas 4.17 and 4.18), we conclude

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{(t,a,b)\in J} \left\{ (|a|\vee|b|)(|\alpha|+|\beta|)\,n^{-\frac{2}{3}} \left| \sum_{i=1}^{n} \Big(\big(\mathbb{1}_{X_i\leq\tau+n^{-\nu}t} - \mathbb{1}_{X_i\leq\tau}\big) \right.\right.\right. \\
&\qquad\qquad\qquad \left.\left.\left. - \mathbb{E}\big(\mathbb{1}_{X_i\leq\tau+n^{-\nu}t} - \mathbb{1}_{X_i\leq\tau}\big)\Big) \right| \geq D\,\|(t,a,b)\|^2 \right\}\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{(t,a,b)\in J} \left\{ (|a|\vee|b|)(|\alpha|+|\beta|)\,n^{\frac{1}{3}} \left| V^{(2)}_{\frac{1}{3},n}(t) \right| \geq D\,\|(t,a,b)\|^2 \right\}\right) \\
&= 0\,,
\end{aligned}$$

for all $d > 0$. □

7.4 Distributional convergence

To conclude this section we will reformulate the result of [BM07, Theorem 2.1].

Theorem 7.8 If (A1), (A2a), (A3) and (C1)-(C2) hold and $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$, as $n \to \infty$, then

$$n^{\frac{1}{3}} \begin{pmatrix} \tau_n - \tau \\ \alpha_n - \alpha \\ \beta_n - \beta \end{pmatrix} \xrightarrow{\mathscr{L}} \lambda \begin{pmatrix} 1 \\ \eta_1 \\ \eta_2 \end{pmatrix},$$

where

$$\lambda = \operatorname*{argmin}_{t\in\mathbb{R}} \big(C_1 B(t) + C_2 t^2 \big)$$

and $(B(t))_{t\in\mathbb{R}}$ is a standard two-sided Brownian Motion,

$$C_1 = \frac{V(\tau)}{\beta-\alpha}\,, \quad C_2 = \frac{1}{2}\left[\frac{m'(\tau)}{\beta-\alpha} - \frac{1}{4}\frac{F'(\tau)}{F(\tau)-F(\tau)^2}\right],$$

and

$$\eta_1 = \frac{(\beta-\alpha)F'(\tau)}{2F(\tau)}\,, \quad \eta_2 = \frac{(\beta-\alpha)F'(\tau)}{2\bar{F}(\tau)}\,.$$

Remark 7.9 Under which mild conditions $(\tau_n, \alpha_n, \beta_n) \xrightarrow{\mathbb{P}} (\tau, \alpha, \beta)$, was shown in Theorem 5.4 and Corollary 5.5. Furthermore, note that the quadratic drift function opens in the correct direction (i.e. $C_2 > 0$, in order to obtain a minimum) if condition (7.1) is fulfilled. Apart from this, note that the not strict version of Inequality (7.1) is a necessary condition for (A1), see Lemma 2.1(vii). In contrast, the authors of [BM07] merely used $m'(\tau) \neq 0$ as an assumption instead. This, however, is not sufficient to prove the statement of Theorem 2.1 in their work. Within their proof the difference $S(t,a,b)-S(\tau,\alpha,\beta)$ (in our notation) is estimated from below by a quadratic function (2nd Equation in their proof of Theorem 2.1). But this estimate is only valid if (7.1) is fulfilled. If this is the case then the Hessian matrix $\mathrm{H}_S(\tau,\alpha,\beta)$ (there denoted as V) is positive definite. The relevant estimate in this work occurs in Lemma 7.7 and is decisive for the form of the drift function of the Brownian motion and the convergence rate. It can be expected that if the regression curve m is too flat at the split point, that is,

$$\frac{m'(\tau)}{\beta-\alpha} = \frac{1}{4}\frac{F'(\tau)}{F(\tau)-F(\tau)^2}\,,$$

the estimator converges at a slower rate than $n^{1/3}$. According to this, the drift function needs to be adapted depending on the estimation from below in Lemma 7.7, or on $S(t,a,b)-S(\tau,\alpha,\beta)$, respectively. Moreover, in what extent this affects the discussion in the last paragraph of page 547 in [BM07] about unstable confidence intervals when the regression curve is too flat at the split point is yet to be clarified.

Finally, we will give a proof of the statement.

Proof of Theorem 7.8. Set $\xi_n := n^{1/3}(\tau_n - \tau, \alpha_n - \alpha, \beta_n - \beta)$ and note that by Lemma 3.9, $\xi_n \in \mathrm{Arginf}(\tilde{Z}_n)$ and $\tilde{Z}_n \in D(\mathbb{R}^3)$. Using the results of Theorem 7.2 and 7.6 together with Proposition 3.12 implies

$$\limsup_{n\to\infty} \mathbb{P}(\xi_n \in F) \le \mathbb{P}\big(\mathrm{Arginf}(\tilde{Z}) \cap F \neq \emptyset\big) \tag{7.15}$$

for all closed sets $F \subseteq \mathbb{R}^3$, where $\tilde{Z}(t)$ is the process from Definition 7.1. Since $B(t)$ is a sample-continuous process, $\tilde{Z}(t)$ is as well and, hence, $\mathrm{Arginf}(\tilde{Z}) = \mathrm{Argmin}(\tilde{Z})$ almost surely by [Fer15, Lemma 2.1 and Lemma 2.2(1)]. Following the rules regarding the calculation of infima, see [RW98, Propositions 1.35-1.37], we have for each $t \in \mathbb{R}$

$$\begin{aligned}
&\inf_{(a,b)\in\mathbb{R}^2} \tilde{Z}(t,a,b)\\
&= 2(\beta-\alpha)\tilde{Z}^{(1)}(t) + \inf_{(a,b)\in\mathbb{R}^2} \tilde{Z}^{(2)}(t,a,b)\\
&= 2(\beta-\alpha)V(\tau)F'(\tau)B(t) + (\beta-\alpha)m'(\tau)F'(\tau)t^2\\
&\qquad\qquad + \inf_{(a,b)\in\mathbb{R}^2}\big((\alpha-\beta)F'(\tau)(a+b)t + a^2F(\tau) + b^2\bar{F}(\tau)\big)\\
&= 2(\beta-\alpha)^2F'(\tau)\left(\frac{V(\tau)}{\beta-\alpha}B(t) + \frac{1}{2}\left[\frac{m'(\tau)}{\beta-\alpha} - \frac{1}{4}\frac{F'(\tau)}{F(\tau)-F(\tau)^2}\right]t^2\right).
\end{aligned}$$

For each $(\theta_1,\theta_2,\theta_3) \in \mathrm{Argmin}(\tilde{Z})$ use the characterization of Argmin-sets in [RW98, Propositions 1.35] to realize that

$$\theta_1 \in \operatorname*{Argmin}_{t\in\mathbb{R}}\left\{2(\beta-\alpha)^2F'(\tau)\left(\frac{V(\tau)}{\beta-\alpha}B(t) + \frac{1}{2}\left[\frac{m'(\tau)}{\beta-\alpha} - \frac{1}{4}\frac{F'(\tau)}{F(\tau)-F(\tau)^2}\right]t^2\right)\right\}$$

and

$$\theta_2 = \frac{(\beta-\alpha)F'(\tau)}{2F(\tau)}\theta_1, \quad \theta_3 = \frac{(\beta-\alpha)F'(\tau)}{2\bar{F}(\tau)}\theta_1.$$

Since the maximum of Brownian motion with quadratic drift is attained at a unique point, cf. [JLM10, p. 1894], we find that θ_1 is unique and we are able to specify Equation (7.15) by

$$\limsup_{n\to\infty} \mathbb{P}(\xi_n \in F) \le \mathbb{P}((\theta_1,\theta_2,\theta_3) \in F)$$

for all closed sets $F \subseteq \mathbb{R}^3$, which by Portmanteau's theorem ([Kal97, Theorem 3.25]) implies our claim. □

A Appendix

In this chapter we collect well known results from the literature and state and proof where necessary results and inequalities which would otherwise worsen the readability of this thesis.

A.1 (Stute and Wang) Let $(X_i, Y_i)_{1\le i\le n}$, for $n \in \mathbb{N}$, be i. i. d. copies of a random vector (X, Y) with some bivariate distribution. The vector $\left(X_{i:n}, Y_{[i:n]}\right)$ denote the i–th order statistic $X_{i:n}$ and corresponding concomitant $Y_{[i:n]}$, means $X_{1:n} \le \ldots \le X_{n:n}$ and $Y_{[i:n]}$ is the random variable associated with $X_{i:n}$, $\underline{X}_n := (X_{1:n}, \ldots, X_{n:n})$ and $\underline{Y}_n := \left(Y_{[1:n]}, \ldots, Y_{[n:n]}\right)$. Then

(a) the concomitants are conditionally independent given $\underline{X}_n$:

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\left\{Y_{[i:n]} \le y_i\right\} \,|\, \underline{X}_n = \underline{x}_n\right) = \prod_{i=1}^{n} \mathbb{P}\left(Y_{[i:n]} \le y_i \,|\, \underline{X}_n = \underline{x}_n\right)$$

for all $\underline{y}_n \in \mathbb{R}^n$ and for $\mathbb{P} \circ \underline{X}_n^{-1}$-almost all $\underline{x}_n \in \mathbb{R}^n$,

(b) for each $1 \le i \le n$

$$\mathbb{P}\left(Y_{[i:n]} \le y \,|\, \underline{X}_n = \underline{x}_n\right) = \mathbb{P}\left(Y_{[i:n]} \le y \,|\, X_{i:n} = x_i\right)$$

for all $y \in \mathbb{R}$ and for $\mathbb{P} \circ \underline{X}_n^{-1}$-almost all $\underline{x}_n \in \mathbb{R}^n$,

(c) for each $1 \le i \le n$

$$\mathbb{P}\left(Y_{[i:n]} \le y \,|\, X_{i:n} = x\right) = \mathbb{P}(Y \le y \,|\, X = x)$$

for all $y \in \mathbb{R}$ and for $\mathbb{P} \circ X_{i:n}^{-1}$-almost all $x \in \mathbb{R}$.

Proof. See [SW93, Lemma 2.1]. □

A.2 (Hájek and Rényi inequality) Let $(X_i)_{1\le i\le n}$, for $n \in \mathbb{N}$, be a n-tuple of independent, square integrable and zero-mean random variables and let $b_m \ge b_{m+1} \ge \ldots \ge b_n > 0$ for some $1 \leq m \leq n$. Then for all $\lambda > 0$

$$\mathbb{P}\left(\max_{m\leq k\leq n} b_k \left|\sum_{i=1}^{k} X_i\right| \geq \lambda\right) \leq \lambda^{-2}\left\{b_m^2 \sum_{i=1}^{m} \mathrm{Var}(X_i) + \sum_{i=m+1}^{n} b_i^2 \mathrm{Var}(X_i)\right\} .$$

Proof. The proof for $m = 1$ is stated in [CT97, Theorem 7.4.8]. We can show the general case by shifting the summation index $b'_k := b_{m+k-1}$, $X'_1 := \sum_{i=1}^{m} X_i$, $X'_k := X_{m+k-1}$ for $1 < k \leq n'$ and $n' := n - m + 1$. The tuple $(X'_i)_{1\leq i\leq n'}$ is still independent and zero-mean. Thus, by Hájek and Rényi inequality for $m = 1$

$$\begin{aligned}
&\mathbb{P}\left(\max_{m\leq k\leq n} b_k \left|\sum_{i=1}^{k} X_i\right| \geq \lambda\right) \\
&= \mathbb{P}\left(\max_{1\leq k\leq n'} b'_k \left|\sum_{i=1}^{k} X'_i\right| \geq \lambda\right) \\
&\leq \lambda^{-2} \sum_{i=1}^{n'} b_i'^2 \mathrm{Var}(X'_i) \\
&= \lambda^{-2}\left\{b_m^2 \sum_{i=1}^{m} \mathrm{Var}(X_i) + \sum_{i=m+1}^{n} b_i^2 \mathrm{Var}(X_i)\right\} .
\end{aligned}$$

□

A.3 (Hájek and Rényi inequality for martingale difference sequences) Let $(S_k, \mathscr{F}_k)_{1\leq k\leq n}$, for $n \in \mathbb{N}$, be a discrete-time $\mathscr{L}^2$-martingale and let $b_m \geq b_{m+1} \geq \ldots \geq b_n > 0$ for some $1 \leq m \leq n$, $S_0 := 0$. Then for all $\lambda > 0$

$$\mathbb{P}\left(\max_{m\leq k\leq n} b_k |S_k| \geq \lambda\right) \leq \lambda^{-2}\left\{b_m^2 \sum_{k=1}^{m} \mathrm{Var}(S_k - S_{k-1}) + \sum_{k=m+1}^{n} b_k^2 \mathrm{Var}(S_k - S_{k-1})\right\} .$$

Proof. For the case $m = 1$ we refer to [GS77, Korollar 6.6.4]. For arbitrary $m \in \mathbb{N}$, the result is formulated as an exercise in [Excercise A.10.3,[SW86]]. A generalization may be conducted analogue to the proof of A.2. □

A.4 (Hájek and Rényi inequality for backwards martingale difference sequences) Let $(S_k, \mathscr{F}_k)_{1\leq k\leq n}$, for $n \in \mathbb{N}$, be a discrete-time $\mathscr{L}^2$-backwards martingale and $b_m \geq b_{m-1} \geq \ldots \geq b_1 > 0$ for some $1 \leq m \leq n$, $S_{n+1} := 0$. Then for all $\lambda > 0$

$$\mathbb{P}\left(\max_{1\leq k\leq m} b_k |S_k| \geq \lambda\right) \leq \lambda^{-2}\left\{b_m^2 \sum_{k=m}^{n} \mathrm{Var}(S_k - S_{k+1}) + \sum_{k=1}^{m-1} b_k^2 \mathrm{Var}(S_k - S_{k+1})\right\} .$$

Proof. Define $S_k^* := S_{n-k+1}$, $\mathscr{F}_k^* := \mathscr{F}_{n-k+1}$ and $b_i^* := b_{n-k+1}$ and note that $\left(S_k^*, \mathscr{F}_k^*\right)_{1\leq k\leq n}$ is a $\mathscr{L}^2$-martingale and $b_{n-m+1}^* \geq \ldots b_{n-1}^* \geq b_n^* > 0$. Thus, we can use A.3 and get

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq k\leq m} b_k\,|S_k| \geq \lambda\right)\\
&= \mathbb{P}\left(\max_{n-m+1\leq j\leq n} b_j^*\left|S_j^*\right| \geq \lambda\right)\\
&\leq \lambda^{-2}\left\{\left(b_{n-m+1}^*\right)^2 \sum_{k=1}^{n-m+1} \operatorname{Var}(S_k^* - S_{k-1}^*) + \sum_{k=n-m+2}^{n} \left(b_k^*\right)^2 \operatorname{Var}(S_k^* - S_{k-1}^*)\right\}\\
&= \lambda^{-2}\left\{b_m^2 \sum_{k=m}^{n} \operatorname{Var}(S_k - S_{k+1}) + \sum_{k=1}^{m-1} b_k^2 \operatorname{Var}(S_k - S_{k+1})\right\}\,. \qquad \square
\end{aligned}$$

A.5 Let $F : \mathbb{R} \to \mathbb{R}$ be a càdlàg function which is strictly increasing within an ε-neighborhood $U_\varepsilon(a)$ of $a \in \mathbb{R}$ and let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function with $\int_{(a-\varepsilon,a+\varepsilon]} g\,\mathrm{d}F < \infty$. If g is right-continuous in a then

$$\lim_{v\downarrow 0} \frac{\int_{(a,a+v]} g(x)F(\mathrm{d}x)}{F(a+v) - F(a)} = g(a+)\,.$$

If otherwise g is left-continuous in a then

$$\lim_{v\downarrow 0} \frac{\int_{(a-v,a)} g(x)F(\mathrm{d}x)}{F(a-) - F(a-v)} = g(a-)\,.$$

Proof. Let $(v_n)_{n\in\mathbb{N}}$ be a sequence with $v_n \downarrow 0$. For each $n \in \mathbb{N}$ set $k_n := \inf_{x\in(a,a+v_n]} g(x)$ and $K_n := \sup_{x\in(a,a+v_n]} g(x)$. Then $(k_n)_{n\in\mathbb{N}}$ is monotonically increasing and $(K_n)_{n\in\mathbb{N}}$ monotonically decreasing with

$$k_n\left[F(a+v_n) - F(a)\right] \leq \int_{(a,a+v_n]} g(x)F(\mathrm{d}x) \leq K_n\left[F(a+v_n) - F(a)\right]\,.$$

By the monotonicity of $(k_n)_{n\in\mathbb{N}}$ and $(K_n)_{n\in\mathbb{N}}$ we have that

$$\lim_{n\to\infty} k_n = \lim_{n\to\infty} \inf_{x\in(a,a+v_n]} g(x) = \sup_{n\in\mathbb{N}} \inf_{x\in(a,a+v_n]} g(x) = \liminf_{x\downarrow a} g(x)$$

and

$$\lim_{n\to\infty} K_n = \lim_{n\to\infty} \sup_{x\in(a,a+v_n]} g(x) = \inf_{n\in\mathbb{N}} \sup_{x\in(a,a+v_n]} g(x) = \limsup_{x\downarrow a} g(x)\,.$$

Since g is right-continuous in a we get $\liminf_{x\downarrow a} g(x) = \limsup_{x\downarrow a} g(x) = g(a+)$ and, hence,

$$\lim_{n\to\infty} \frac{\int\limits_{(a,a+v_n]} g(x)F(\mathrm{d}x)}{F(a+v_n)-F(a)} = g(a+)\,.$$

The other case is quite similar. Set $\tilde{k}_n := \inf_{x\in(a-v_n,a)} g(x)$ and $\tilde{K}_n := \sup_{x\in(a-v_n,a)} g(x)$ and note that

$$\lim_{n\to\infty} \tilde{k}_n = \liminf_{x\uparrow a} g(x) \quad \text{and} \quad \lim_{n\to\infty} \tilde{K}_n = \limsup_{x\uparrow a} g(x)\,. \qquad \square$$

A.6 Let Q be a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and $E \in \mathscr{B}(\mathbb{R})$ with $Q(E) = 1$. If $\tau \in \mathbb{R}$ and there exists an $\varepsilon > 0$ such that $Q((\tau, \tau + r]) > 0$ for all $r \in (0, \varepsilon)$ then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in E$ for sufficiently large $n \in \mathbb{N}$ and $x_n \searrow \tau$. If, otherwise, for $\tau \in \mathbb{R}$ there exists an $\varepsilon > 0$ such that $Q((\tau - r, \tau)) > 0$ for all $r \in (0, \varepsilon)$ then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in E$ for sufficiently large $n \in \mathbb{N}$ and $x_n \nearrow \tau$.

Proof. We will show that $(\tau, \tau + r] \cap E \neq \emptyset$ for all $r \in (0, \varepsilon)$. Conversely assume that there is an $r \in (0, \varepsilon)$ such that $(\tau, \tau + r] \cap E = \emptyset$. Then $(\tau, \tau + r] \subseteq E^{\mathrm{C}}$ and

$$0 = Q(E^{\mathrm{C}}) \geq Q((\tau, \tau + r]) > 0\,,$$

a contradiction. Let $(r_n)_{n\in\mathbb{N}} \subseteq (0, \varepsilon)$ with $r_n \downarrow 0$ then for sufficiently large $n \in \mathbb{N}$ there exists $x_n \in (\tau, \tau + r_n] \cap E$ and $x_n \searrow \tau$. With similar arguments find $(\tilde{x}_n)_{n\in\mathbb{N}} \subseteq E$ with $\tilde{x}_n \nearrow \tau$. $\square$

A.7 Let $[a, b]$ be an interval in $\mathbb{R}$ and g be a càdlàg function on $[a, b]$, then

$$\sup_{a<x\leq b} g(x-) = \sup_{a\leq x<b} g(x)\,.$$

Proof. Define $A := \sup_{a<x\leq b} g(x-)$ and $B := \sup_{a\leq x<b} g(x)$. Due to the definition of

$g(x-)$ for all $x \in (a,b]$ and $(x_n)_{n\in\mathbb{N}} \subseteq [a,b)$ with $x_n \uparrow x$

$$g(x-) = \lim_{n\to\infty} g(x_n) \leq \sup_{a\leq x<b} g(x)\,,$$

and therefore $A \leq B$. Let $B = \lim_{n\to\infty} g(x_n)$ for a sequence $(x_n)_{n\in\mathbb{N}} \subseteq [a,b)$. Since $[a,b]$ is compact, we have $x \in [a,b]$. For each $n \in \mathbb{N}$ either $x_n \in [a,x)$ or $x_n \in [x,b)$. This implies the existence of a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ which is either monotonically decreasing or strictly monotonically increasing. If $x_{n_k} \uparrow x$ strictly, then $x \in (a,b]$ and

$$B = \lim_{x_{n_k}\uparrow x} g(x_{n_k}) = g(x-) \leq \sup_{a<x\leq b} g(x-) = A\,.$$

In case of $x_{n_k} \downarrow x$ we have $x \in [a,b)$ and since g is right-continuous there exists a sequence $\big(y_{n_k}\big)_{k\in\mathbb{N}}$ with $y_{n_k} \in (x_{n_k}, b]$ and $\big|g(x_{n_k}) - g(y)\big| \leq 1/k$ for all $y \in [x_{n_k}, y_{n_k}]$. Hence, $g(y_{n_k}-) \geq g(x_{n_k}) - 1/k$ and

$$B = \lim_{k\to\infty} \big(g(x_{n_k}) - 1/k\big) \leq \lim_{k\to\infty} g(y_{n_k}-) \leq A\,. \qquad \square$$

List of Acronyms and Abbreviations

càdlàg right-continuous with left limits
càglàd left continuous with right limits
CLT Central Limit Theorem
CMT Continuous mapping Theorem

fidis finite-dimensional marginal distributions

i. i. d. independent and identically distributed

SLLN Strong Law of Large Numbers

List of Symbols

Probability theory

$(\Omega, \mathscr{A}, \mathbb{P})$	probability space
$\mathscr{B}(A)$	Borel σ-Algebra on the set A
$\sigma(X)$	σ-Algebra generated by the random variable X
$\mu \otimes \nu$	product of the measures μ and ν
$\mathbb{P}(A)$	probability of the set A
$\mathbb{P}_X = \mathbb{P} \circ X^{-1}$	distribution of the random variable X
$\mathbb{P}_{(X,Y)}$	joint distribution of X and Y
$\mathbb{P}_{Y\|X}(x, \cdot)$	conditional distribution of the random variable Y given $X = x$
$\mathbb{E}(X)$	expectation of the random variable X
$\mathbb{E}(Y\|X)$	conditional expectation of the random variable Y given $\sigma(X)$
$\mathbb{E}(Y\|X = x)$	conditional expectation of the random variable Y given $X = x$
$f_{Y\|X}(x, y)$	conditional probability density function of Y given $X = x$
φ_X	characteristic function of the random variable X
$\sim$	distributed as
$\stackrel{\mathscr{L}}{=}$	equal in distribution
$\delta_x(\cdot)$	Dirac's measure at the point x
$\mathscr{N}\left(\mu, \sigma^2\right)$	Normal distribution with mean μ and variance σ^2
$\xrightarrow{\mathscr{L}}$	convergence in distribution
$\xrightarrow{w}$	weak convergence
$\xrightarrow{\mathbb{P}}$	convergence in probability
(X_i, Y_i)	i-th sample of the random variable (X, Y)
$\left(X_{i:n}, Y_{[i:n]}\right)$	the i-th order statistic and its corresponding concomitant $Y_{[i:n]}$
$\underline{X}_n = (X_{1:n}, \ldots, X_{n:n})$	vector of order statistics of $X_1, \ldots, X_n$
$\underline{Y}_n = (Y_{[1:n]}, \ldots, Y_{[n:n]})$	vector of concomitants of $\underline{X}_n$

Analysis

A^{C}	complement of the set A
$\mathbb{1}_A(\cdot)$	indicator function of the set A
$a \wedge b$	minimum of a and b
$a \vee b$	maximum of a and b
(a,b), $(a,b]$, $[a,b]$	open, half open, closed intervals in $\mathbb{R}$
$\langle a, b\rangle$	scalar product in $\mathbb{R}^d$
$a_n \uparrow a$ $(a_n \downarrow a)$	a_n converges monotonically to a from below (from above)
$a_n \nearrow a$ $(a_n \searrow a)$	a_n converges to a with $a_n < a$ (or $a_n > a$) for each $n \in \mathbb{N}$
$f(a-) = f_-(a) = \lim\limits_{x \nearrow a} f(x)$	left-hand limit
$f(a+) = f_+(a) = \lim\limits_{x \searrow a} f(x)$	right-hand limit
f' , f''	first and second derivative of the function f
f'_- (f'_+)	left-hand (right-hand) derivative of f
$\partial_x f$	partial derivative of f with respect to x
$\nabla f = \left(\partial_{x_1} f, \ldots, \partial_{x_d} f\right)^\top$	gradient of f
$\mathrm{H}_f = \left(\partial_{x_i} \partial_{y_j} f\right)_{i,j \in \{1,\ldots,d\}}$	Hessian matrix of f
$\mathrm{diag}(\gamma_1, \ldots, \gamma_d)$	diagonalmatrix with the entries $\gamma_1, \ldots, \gamma_d$ on the main diagonal
$U_\varepsilon(\tau)$	ε-neighborhood of τ
$\lVert\cdot\rVert$	Euclidean norm
$\lVert\cdot\rVert_\infty$	maximum norm
$\lVert f \rVert_a^b = \sup_{a \le t \le b} f(t)$	supremum of f on the intervall $[a,b]$
$\pi_S(f)$	projection map

Bibliography

[Als98] Gerold Alsmeyer. *Wahrscheinlichkeitstheorie*. Skripten zur Mathematischen Statistik, Nr. 30. WWU Münster, 1998.

[App04] David Applebaum. *Lévy processes and stochastic calculus*. Vol. 93. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2004.

[Bil68] Patrick Billingsley. *Convergence of probability measures*. New York: Wiley, 1968.

[Bil95] Patrick Billingsley. *Probability and measure*. 3. ed. Wiley, 1995.

[Bil99] Patrick Billingsley. *Convergence of probability and measures*. 2nd ed. John Wiley & Sons Inc., New York, 1999.

[BM07] M. Banerjee and I. W. McKeague. "Confidence sets for split points in decision trees". In: *The Annals of Statistics* 35(2) (2007), pp. 543–574.

[Bre+84] Leo Breiman et al. *Classification and regression trees*. Chapman and Hall/CRC, 1984.

[Bre92] Leo Breiman. *Probability*. Unabridged, corr. republication of the work 1. publ. 1968. Philadelphia : SIAM, 1992.

[Cha55] Li-Chien Chang. "On the ratio of an empirical distribution function to the theoretical distribution function". In: *Acta Mathematica Sinica. Shuxue Xuebao* 5 (1955), pp. 347–368.

[CT97] Y. S. Chow and H. Teicher. *Probability theory*. 3rd ed. Springer, New York, 1997.

[Fer04] D. Ferger. "A continuous mapping theorem for the argmax-functional in the non unique case". In: *Statistica Neerlandica* (2004).

[Fer09] D. Ferger. "Stochastische Prozesse mit Strukturbrüchen". In: *Dresdner Schriften zur Mathematischen Stochastik* (2009).

[Fer15] D. Ferger. “Arginf-sets of multivariate cadlag processes and their convergence in hyperspace topologies”. In: *Theory of Stochastic Processes* 20.2 (2015), pp. 13–41.

[Fis02] Gerd Fischer. *Lineare Algebra: Eine Einführung für Studienanfänger*. 13., durchges. Aufl. Braunschweig, Wiesbaden: Vieweg, 2002.

[FK09] Dietmar Ferger and Jens Klotsche. “Estimation of split-points in binary regression”. In: *Statist. Decisions* 27.2 (2009), pp. 93–128.

[For08] Otto Forster. *Analysis 2 Differentialrechnung im R^n, gewöhnliche Differentialgleichungen*. 8., aktualisierte Aufl. Wiesbaden: Vieweg + Teubner, 2008.

[FV09] D. Ferger and D. Vogel. “Weak convergence of the empirical process and the rescaled empirical distribution function in the Skorokhod product space”. In: *Teor. Veroyatn. Primen.* 54.4 (2009), pp. 750–770.

[Geo13] Hans-Otto Georgii. *Stochastics*. De Gruyter Textbook. Introduction to probability and statistics, Translated from the German original by Marcel Ortgiese, Ellen Baake and Georgii. Walter de Gruyter & Co., Berlin, 2013.

[GS77] P. Gänssler and W. Stute. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 1977.

[Heu91] Harro Heuser. *Lehrbuch der Analysis 2*. 6., überarb. Aufl. Stuttgart: Teubner, 1991.

[HS75] E. Hewitt and K. Stromberg. *Real and abstract analysis : a modern treatment of the theory of functions of a real variable*. 3. print. Springer, 1975.

[JLM10] Svante Janson, Guy Louchard, and Anders Martin-Löf. “The maximum of Brownian motion with parabolic drift”. In: *Electronic Journal of Probability* 15 (2010), no. 61, 1893–1929.

[JS03] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*. Second. Vol. 288. Springer-Verlag, Berlin, 2003.

[Kal97] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, 1997.

[Kle06] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2006.

[Kos08] Michael R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York, 2008.

[Kou02] Hira L. Koul. *Weighted empirical processes in dynamic nonlinear models*. 2. ed. New York, Berlin, Heidelberg: Springer, 2002.

[ŁR86] Zbigniew A. Łagodowski and Zdzisław Rychlik. "Weak convergence of probability measures on the function space $D_d[0, \infty)$". In: *Bull. Polish Acad. Sci. Math.* 34.5-6 (1986), pp. 329–335.

[RK10] J. Rentsch and R. Kropp. "Schätzung von Splitpoints in Entscheidungsbäumen". MA thesis. Technische Universität Dresden, 2010.

[Rud76] Walter Rudin. *Principles of mathematical analysis*. Third. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976.

[RW98] R. T. Rockafellar and R. J-B. Wets. *Variational Analysis*. Springer, Berlin, 1998.

[SBK16] Rui Song, Moulinath Banerjee, and Michael R. Kosorok. "Asymptotics for change-point models under varying degrees of mis-specification". In: *The Annals of Statistics* 44.1 (2016), pp. 153–182.

[Sei12] Emilio Seijo. "Statistical inference in two non-standard regression problems". PhD thesis. Columbia University in the City of New York, 2012.

[Sie96] Gerard Sierksma. *Linear and integer programming theory and practice*. New York: Dekker, 1996.

[SP12] René L. Schilling and Lothar Partzsch. *Brownian motion - an introduction to stochastic processes*. De Gruyter, 2012.

[SS11a] Emilio Seijo and Bodhisattva Sen. "A continuous mapping theorem for the smallest argmax functional". In: *Electronic Journal of Statistics* 5 (2011), pp. 421–439.

[SS11b] Emilio Seijo and Bodhisattva Sen. "Change-point in stochastic design regression and the bootstrap". In: *The Annals of Statistics* 39.3 (2011), pp. 1580–1607.

[Stu97] Winfried Stute. "Nonparametric Model Checks for Regression". In: *The Annals of Statistics* 25.2 (1997), pp. 613–641.

[SW86] Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics*. Wiley, 1986.

[SW93] W. Stute and J.-L. Wang. "The Strong Law under Random Censorship". English. In: *The Annals of Statistics* 21.3 (1993), pp. 1591–1607.

[Til07] D. Tillich. “Schätzung der Sprungstelle bei einstufigen Regressionsfunktionen”. MA thesis. Technische Universität Dresden, 2007.

[VW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. With applications to statistics. Springer-Verlag, New York, 1996.

[Wit85] H. Witting. *Mathematische Statistik: Parametrische Verfahren bei festem Stichprobenumfang*. Teubner B.G. GmbH, 1985.

[WM95] Hermann Witting and Ulrich Müller-Funk. *Mathematische Statistik. 2, Asymptotische Statistik: parametrische Modelle und nicht-parametrische Funktionale*. Stuttgart: Teubner, 1995.

Acknowledgments

First of all I wish to express my sincere gratitude to my supervisor Prof. Dr. Dietmar Ferger, for his ongoing support, mathematical guidance and many fruitful discussions throughout my research.

I would like to thank the second assessor Prof. Dr. Dr. h. c. Winfried Stute for his work. In the same manner I would like to thank the members of the Doctorate Commission for their time and effort. I would also like to thank my colleagues for the enjoyable cooperation and the friendly atmosphere at the Institute of Mathematical Stochastics at the Technische Universität Dresden.

Many thanks are owed to Jan for many mathematical discussions and helpful suggestions. I am also grateful to my family and friends for their important backing and support. Finally, my warm thanks go to Thomas for his immeasurable love, support and motivation to keep on going.

Affirmation

I herewith declare that I have produced this thesis without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This thesis has not previously been presented in identical or similar form to any other German or foreign examination board.
The present thesis has been produced at the Institute of Mathematical Stochastics, Faculty of Mathematics, School of Science, Technische Universität Dresden under the scientific supervision of Prof. Dr. Dietmar Ferger.
There have been no prior attempts to obtain a PhD degree at any university.
I accept the Doctorate Regulations of the School of Science of the Technische Universität Dresden, issued 23rd February 2011.

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.
Die vorliegende Arbeit wurde am Institut für Mathematische Stochastik, Fakultät Mathematik, Bereich Mathematik und Naturwissenschaften, Technische Universität Dresden unter der wissenschaftlichen Betreuung von Prof. Dr. Dietmar Ferger angefertigt.
Es haben bisher keine früheren Promotionsverfahren an Universitäten stattgefunden.
Ich erkenne die Promotionsordnung des Bereiches Mathematik und Naturwissenschaften der Technische Universität Dresden vom 23. Februar 2011 an.

Dresden, 29.04.2020